

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 08:35:50 ; Search time 66 Seconds  
(without alignments)  
205.933 Million cell updates/sec

Title: US-09-882-434A-1  
Perfect score: 551  
Sequence: 1 MASTKLFPSVTVMMLAMA.....FGSARACNPGKMSITQC 102.

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.GeneSeq 101002:\*

1:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23:	/SID27/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	551	100.0	102	18	AAW31738
2	74.5	13.5	486	21	AAAB10456
3	71	12.9	1280	22	AAAB1064
4	71	12.9	1283	22	AAAB1065
5	70.5	12.8	267	22	AAAB6131
6	70.5	12.8	267	22	AAAB5431
7	70.5	12.8	267	22	AAU17539
8	70.5	12.8	524	19	AAW34998
9	69	12.5	138	22	AAU50762
10	69	12.5	1280	22	AAAB1068

11	69	12.5	1280	22	AAE00304	Dog P-glycoprotein
12	69	12.5	1281	22	AAE00303	Dog (FGP) P-glycop
13	69	12.5	1281	22	AAE00308	Dog P-glycoprotein
14	69	12.5	1281	22	AAE00309	Dog P-glycoprotein
15	69	12.5	1281	22	AAE00310	Dog P-glycoprotein
16	68.5	12.4	1280	19	AAW48398	N-terminal single
17	68.5	12.4	1280	19	AAW49000	Human P glycoprote
18	68	12.3	1280	8	AAE70452	Sequence encoded b
19	68	12.3	1280	11	AAE04868	Protein encoded by
20	68	12.3	1280	14	AAE44297	Sequence encoded b
21	68	12.3	1280	15	AAE63524	Human P-glycoprote
22	68	12.3	1280	16	AAW44073	Human P-glycoprote
23	68	12.3	1280	19	AAW48397	Human multiding re
24	68	12.3	1280	19	AAW48399	wild-type human p
25	68	12.3	1280	21	AAE58186	C-terminal single
26	68	12.3	1280	21	AAE58187	Human wild-type mu
27	68	12.3	1280	22	AAU04347	Human G155V mutant
28	68	12.3	1280	22	AAE81066	Human multiding re
29	68	12.3	1280	22	AAE81959	Human P-glycoprote
30	68	12.3	1280	22	AAE00306	Human MDRI. Homo
31	68	12.3	1280	23	AAE24211	Human P-glycoprote
32	68	12.3	1280	23	AAU97493	Human MDR protein.
33	68	12.3	1280	23	AAE18964	Human P-glycoprote
34	68	12.3	1280	23	AAE18965	Human wild type p-
35	68	12.3	1280	23	AAE18966	Human P-gp mutant,
36	68	12.3	1280	23	AAE18967	Human P-gp mutant,
37	68	12.3	1280	23	AAE18968	Human P-gp mutant,
38	68	12.3	1280	23	AAE18969	Human P-gp mutant,
39	68	12.3	1280	23	AAE18970	Human P-gp mutant,
40	68	12.3	1280	23	AAE18971	Human P-gp mutant,
41	68	12.3	1280	23	AAE18972	Human P-gp mutant,
42	68	12.3	1280	23	AAE18973	Human P-gp mutant,
43	68	12.3	1280	23	AAE18974	Human P-gp mutant,
44	68	12.3	1280	23	AAE18975	Human P-gp mutant,
45	68	12.3	1280	23	AAE18976	Human P-gp mutant,

## ALIGNMENTS

RESULT 1					
AAW31738					
ID	AAW31738	standard	Protein	102 AA.	
XX					
AC	AAW31738;				
XX					
DT	27-MAR-1998	(first entry)			
DE	Macadamia integrifolia antimicrobial protein 1.				
XX					
XX	Antimicrobial protein 1; growth inhibition; fungus; bacterium;				
KW	fungi; bacteria; pathogen control.				
KM					
XX					
OS	Macadamia integrifolia.				
XX					
FH	Key	Location/Qualifiers			
FT	Peptide	1..26			
FT		/label= sig_peptide			
FT	Peptide	27..102			
FT		/label= mat_peptide			
XX					
XX					
PN	WO9728185-A1.				
XX					
PD	07-AUG-1997.				
XX					
PF	31-JAN-1997;	97WO-AU00052.			
XX					
PR	31-JAN-1996;	96AU-0007802.			
XX					
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.				
XX					
PI	Goulter KC, Green JL, Harrison SJ, Manners JM, Marcus JP;				
XX					

DR WPI; 1997-448317/41.  
DR N-PSDB; AAT88851.  
XX  
PT New anti-microbial protein from *Macadamia integrifolia* - active in  
PT inhibiting the growth of fungi and bacteria in plants and animals  
XX  
PS Claim 1; Pages 23-24; 38pp; English.  
XX  
CC The present sequence is *Macadamia integrifolia* antimicrobial  
CC protein 1, which exerts inhibitory activity on the growth of fungi  
CC and bacteria. It can be used for the control of pathogens in plants  
CC and animals.  
XX  
SQ Sequence 102 AA;  
Query Match 100.0%; Score 551; DB 18; Length 102;  
Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MASTKLFSSVITVNMILAMASEMVGSAFTVWSGPGCNRAERYSKGCCSAIHQKGGYDF 60  
DB 1 MASTKLFSSVITVNMILAMASEMVGSAFTVWSGPGCNRAERYSKGCCSAIHQKGGYDF 60  
QY 61 SYTGTAALYNQAGCSGVAHTRFGSSARACNPFQWKSGIFIQ 102  
DB 61 SYTGTAALYNQAGCSGVAHTRFGSSARACNPFQWKSGIFIQ 102  
RESULT 2  
AAB10456  
ID AAB10456 standard; Protein; 486 AA.  
XX  
AC AAB10456;  
XX  
DT 11-DEC-2000 (first entry)  
XX  
DE S. pastorianus mutant hexokinase protein.  
XX  
KW Hexokinase; mutant; creatinine kinase detection; glucose detection.  
XX  
OS Saccharomyces pastorianus.  
XX  
PN JP2000175688-A.  
XX  
PD 27-JUN-2000.  
XX  
PF 17-DEC-1998; 98JP-0359018.  
XX  
PR 17-DEC-1998; 98JP-0359018.  
XX  
PS (TOYM) TOYOSO KK.  
XX  
WPI; 2000-501189/45.  
DR N-PSDB; AAA71483.  
XX  
PT Mutant hexokinase, useful for the detection of creatinine kinase and  
PT glucose, comprises a deletion, addition or insertion in the wild-type  
PT amino acid sequence -  
XX  
PS Claim 3; Page 8-9; 13pp; Japanese.  
XX  
CC This invention describes a novel mutant hexokinase (I) comprising a  
CC protein having hexokinase activity in which at least one amino acid  
CC is added, deleted, inserted or replaced. The amino acid sequence  
CC constitutes the protein having hexokinase activity which remains after  
CC the protein is treated at 50 degrees C for 30 minutes in liquid state,  
CC and which is higher than that of the protein before it is mutated. (I)  
CC is used for the detection of creatinine kinase and glucose. This sequence  
CC represents the *Saccharomyces pastorianus* hexokinase protein which is  
CC described in the method of the invention.  
XX  
SQ Sequence 486 AA;

Query Match 13.5%; Score 74.5; DB 21; Length 486;  
Best Local Similarity 29.2%; Pred. No. 5.5;  
Matches 28; Conservative 10; Mismatches 27; Indels 31; Gaps 6;  
QY 8 FSVITVMM---LIAMASEMVGSAFTVWSGPGCNRAERYSKGCCSAIHQKGGYDFSYTG 64  
DB 369 FGINTTVQERKLIRRLSELIGA-----RAARLSVCGIAAIVCQKRGYK---TG 412  
QY 65 QTAALYNQAGCSGVAHTRF-CSSARACNP-----FGW 95  
DB 413 HIAA-----DGSVYNRYPGFKKAAANALKDIYGW 441  
RESULT 3  
AAB81064  
ID AAB81064 standard; Protein; 1280 AA.  
XX  
AC AAB81064;  
XX  
DT 25-JUN-2001 (first entry)  
XX  
DE Cynomologous monkey P-glycoprotein variant 1.  
XX  
KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;  
KW efflux pump.  
XX  
OS Macaca fascicularis.  
XX  
PN WO200123565-A1.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26592.  
XX  
PR 28-SEP-1999; 99US-0156921.  
XX  
PR 12-OCT-1999; 99US-0158818.  
XX  
PA (GENT-) GENTEST CORP.  
XX  
PI Stocker PJ, Steimel-Crespi DT, Crespi CL;  
DR WPI; 2001-316136/33.  
DR N-PSDB; AAF86127.  
XX  
PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein  
PT (PGP) and homologous PGP polypeptides are useful for predicting  
PT bioavailability of compound and increasing PGP transporter activity in  
PT cell -  
XX  
PS Claim 9; Page 57-59; 84pp; English.  
XX  
CC This invention relates to a polynucleotide sequence encoding a  
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the  
CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member  
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump  
CC exporting small molecules across the cell membrane. The invention  
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence  
CC and protein, and also that of an allelic variant. The PGP polynucleotide  
CC sequence is useful for increasing PGP transporter activity in a cell.  
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport  
CC activity in a mammalian cell. They may also be used for increasing the  
CC bioavailability of a drug. The present sequence represents the  
CC cynomologous monkey P-glycoprotein.  
XX  
SQ Sequence 1280 AA;  
Query Match 12.9%; Score 71; DB 22; Length 1280;  
Best Local Similarity 25.5%; Pred. No. 42;  
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;  
QY 6 LFPSVITVNMILAMASEMV-----NGSAFTVWSGPGCNRAERYSKGCCSAIHQKGGYD 59  
DB 334 VFPSVLIGAFSVGQASPSIEAFANARGAFAEFKIDNKPISDSYSKSGHKPDNKGJLE 393

QY 60 -----FSY-----TGOTALVYNQAGC 75  
DB 394 FRNVHFSYSPRKEVKILKGLNKKVQSGQTVALVGNNGC 434

## RESULT 4

ABAB1065  
ID AAB81065 standard; Protein; 1283 AA.

XX AAB81065;

XX 25-JUN-2001 (first entry)

XX Cynomologous monkey P-glycoprotein variant 2.

XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;  
XX efflux pump.

XX Macaca fascicularis.

XX Key Location/Qualifiers

XX Misc-difference 93..95 /note= "An additional 3 amino acids are present compared  
XX to PGP variant AAB81064"

XX MO200123565-A1.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26592.

XX 28-SEP-1999; 99US-0156921.

XX 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steitel-Crepi DT, Crepi CL;

XX WPI; 2001-316136/33.

XX N-PSDB; AAF86128.

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein  
XX (PGP) and homologous PGP polypeptides are useful for predicting  
XX bioavailability of compound and increasing PGP transporter activity in  
XX cell.

XX Claim 9; Page 65-68; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a  
XX cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the  
XX PGP protein. PGP, also known as multidrug transporter, MDR1 is a member  
XX of the ABC transporter superfamily. The enzyme serves as an efflux pump  
XX exporting small molecules across the cell membrane. The invention  
XX includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence  
XX and protein, and also that of an allelic variant. The PGP polynucleotide  
XX sequence is useful for increasing PGP transporter activity in a cell.  
XX Antisense sequences of the cDNA are useful for inhibiting PGP transport  
XX activity in a mammalian cell. They may also be used for increasing the  
XX bioavailability of a drug. The present sequence represents the  
XX cynomologous monkey P-glycoprotein variant 2. The protein has an  
XX additional 3 amino acids when compared to PGP variant 1 (AAB81065).

XX Sequence 1283 AA;

XX Query Match 12.9%; Score 71; DB 22; Length 1283;

XX Best Local Similarity 25.5%; Pred. No. 42;

XX Matches 25; Conservative 12; Mismatches 33; Indels 26; Gaps 3;

QY 6 LFEVITVMMILIMASENV-----NSAFTVSGPCNNRERYSKGCSAIIHQKGYD 59  
DB 337 VFESVLIGAFVGOASPSIEAFANARGAAFEIFKIDNKPSIDSKSGHKHPINIKNLE 396

QY 60 -----FSY-----TGOTALVYNQAGC 75  
DB 397 FRNVHFSYSPRKEVKILKGLNKKVQSGQTVALVGNNGC 434

## RESULT 5

ABAB96121  
ID ABB96121 standard; Protein; 267 AA.

XX ABB96121;

XX 21-JUN-2002 (first entry)

XX Human testicular antigen SEQ ID NO: 1505.

XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;  
XX reproductive system disorder; urinary system disorder; gene therapy;  
XX cardiovascular disorder; respiratory disorder; neurological disorder;  
XX gastrointestinal disease; infection; cytostatic.

XX Homo sapiens.

XX MO200155317-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01329.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225457.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226868.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227182.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0229513.

XX 06-SEP-2000; 2000US-0230437.

XX 06-SEP-2000; 2000US-0230438.

XX 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231245.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 23-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 12-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-483232/52.  
XX  
XX Nucleic acids encoding 973 human testicular antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating testicular cancer -  
PT  
XX  
XX Claim 11; SEQ ID NO 1505; 766pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 973  
CC human testicular antigens, and fragments of their genomic sequences. The  
CC sequences can be used in the treatment of cardiovascular, urinary system,  
CC reproductive system, immune, respiratory, neurological and  
CC gastrointestinal disorders, infections, and particularly cancer,  
CC especially testicular cancers. The present sequence is a protein of the  
CC invention.  
XX  
XX Sequence 267 AA;  
SQ  
Query Match 12.8%; Score 70.5; DB 22; Length 267;  
Best Local Similarity 27.0%; Pred. No. 7.9;  
Matches 20; Conservative 13; Mismatches 36; Indels 5; Gaps 2;  
QY 28 APTVSGPCNNRARRYSKCGSAHQGGYDFSYTGQTALYNQAGCGVAHTFGSSA 87  
DB 146 SPEFWDSPFNDRGRGH--GLRGAFSAGFGFPAFMEAFSFMNLGCGSGSHITFSSTS 203  
QY 88 ---RACNPFQWKS I 98  
DB 204 FGGSSGSGSGFKSV 217  
RESULT 6  
AAM95431  
ID AAM95431 standard; Protein; 267 AA.  
XX  
XX AAM95431;  
AC AAM95431;  
XX  
XX 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen SEQ ID NO: 4089.  
DE  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX WO200155320-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01339.  
PF

XX 31-JAN-2000; 2000US-0179055.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225256.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 23-AUG-2000; 2000US-0237182.  
 PR 23-AUG-2000; 2000US-0237009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0233401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0235802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250191.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251031.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256179.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM,  
 XX  
 DR WPI; 2001-46570/50.  
 DR N-PSDB; AAL01401.  
 XX  
 PT Isolated nucleic acid molecule encoding a reproductive system antigen

FR 14-AUG-2000; 2000US-0224319.



Query Match 12.8%; Score 70.5; DB 19; Length 524;  
 Best Local Similarity 26.4%; Pred. No. 17;  
 Matches 33; Conservative 12; Mismatches 37; Indels 43; Gaps 8;

QY 3 STKLFFSV-ITVMMILMA-----SEMVNGSAFT--VMSGPG-----CNRR 40  
 DB 63 SMKKFVALFITMFFVSMVAQPAASAAKYSELEGGVIMQAFYDVPGGIWMWDTIRSK 122  
 QY 41 AERYSKGCGCAI-----HKGKG-----YDFSVTGTAAALYNQAGCGVHAHTRFGSS 86  
 DB 123 IPEYEAIGISAIWIPPAKSKMSGYSGYDYPDFDLGE-----YNO---XGIIETRFSGK 175  
 QY 87 ARACN 91  
 DB 176 QELIN 180

RESULT 9  
 ID AAU50762 standard; Protein; 138 AA.  
 AC AAU50762;  
 DT 27-FEB-2002 (first entry)  
 DE Propionibacterium acnes immunogenic protein #11658.  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 OS Propionibacterium acnes.  
 PN W0200181581-A2.  
 XX 01-NOV-2001.  
 XX 20-APR-2001; 2001WO-US12865.  
 XX 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX (CORI-) CORIXA CORP.  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'malsonneuve J, Zhang Y, Jen S, Carter D;  
 XX WPI; 2001-616774/71.  
 DR N-PSDB; AAS59549.  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX Example 1; SEQ ID No 11957; 1069pp; English.  
 PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 138 AA;

Query Match 12.5%; Score 69; DB 22; Length 138;  
 Best Local Similarity 31.5%; Pred. No. 5.4;  
 Matches 29; Conservative 8; Mismatches 25; Indels 30; Gaps 7;

QY 14 MMLTAMASEMVNGSAFTVW-SGPCNNRAERYKCGCSAIHQKGYDFSVTGTAAALYNQ 72  
 DB 53 LMLRLPRESVQ-----WPSGPASIFRA-----SCGTST-----TASTSAAM-V 90

QY 73 AGCGVAHTRFGS-----SARACNPFQWKS 98  
 DB 91 GSCSRIETAGAAASPRISRLASP-GWKS 121

RESULT 10

AA81068

ID AA81068 standard; Protein; 1280 AA.

XX AA81068;

DT 25-JUN-2001 (first entry)

DE Dog P-glycoprotein SEQ ID 7.

XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;  
 KW efflux pump; dog.  
 OS Canis familiaris.  
 PN W0200123565-A1.  
 XX 05-APR-2001.  
 XX 28-SEP-2000; 2000WO-US26592.  
 XX 28-SEP-1999; 99US-0156921.  
 PR 12-OCT-1999; 99US-0158818.  
 XX (GENT-) GENTEST CORP.  
 XX Stocker PJ, Steimel-Crespi DT, Crespi CL;  
 DR WPI; 2001-316136/33.  
 XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein  
 PT (PGP) and homologous PGP polypeptides are useful for predicting  
 PT bioavailability of compound and increasing PGP transporter activity in  
 PT cell -  
 XX Claim 8; Page 73-76; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a  
 CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the  
 CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member  
 CC of the ABC transporter superfamily. The enzyme serves as an efflux pump  
 CC exporting small molecules across the cell membrane. The invention  
 CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence  
 CC and protein, and also that of an allelic variant. The PGP polynucleotide  
 CC sequence is useful for increasing PGP transporter activity in a cell.  
 CC Antisense sequences of the cDNA are useful for inhibiting PGP transport  
 CC activity in a mammalian cell. They may also be used for increasing the  
 CC bioavailability of a drug. The present sequence represents  
 CC P-glycoprotein from a dog.

XX Sequence 1280 AA;





1000

[illegible]

QY 6 LFFSVITVMTLIMASEMV-----NSAFTVWSGPGCNRARERYSKCGSAIHOKGYD 59  
 DB 336 VFESVLIGAFSIGQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGHKKPDNIKMLE 395  
 QY 60 -----FSY-----TGQTALYNQAGC 75  
 DB 396 FKXVHFSPSRKEVKILKGLNLKQSGQTVALVGNSSG 433

RESULT 15  
 ID AAE00310 standard; Protein; 1281 AA.  
 AC AAE00310;  
 DT 13-JUN-2001 (first entry)  
 XX  
 DE Dog P-glycoprotein (PGP) allelic variant (Genotype D).  
 KW Dog; P-glycoprotein allelic variant; PGP, multidrug transporter;  
 KM MDR1; drug bioavailability; transgenic animal; genetic model.  
 XX  
 OS Canis familiaris.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 25 /note= "Asn of Genotypec substituted by Lys"  
 FT Misc-difference 197 /note= "His of Genotypec substituted by Gln"  
 FT Misc-difference 329 /note= "Ser of Genotypec substituted by Thr"  
 FT Misc-difference 1148 /note= "Met of Genotypec substituted by Val"  
 XX  
 PN WQ200123540-A2.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 28-SEP-2000; 2000WO-US26767.  
 XX  
 PR 28-SEP-1999; 99US-0156510.  
 XX  
 PA (GENT-) GENTEST CORP.  
 XX  
 PI Stocker PJ, Steimel-crepi DT, Crespi CL, Reif TC, Patren CJ;  
 DR WPI; 2001-235373/24.  
 XX  
 DR N-PSDB; AAD03506.  
 XX  
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful  
 PT for determining the bioavailability of drugs and for screening for dog  
 PT PGP inhibitors -  
 PS  
 PS Claim 17, Page 108-110, 111pp; English.  
 XX  
 CC The invention relates to dog P-glycoprotein (PGP) also referred  
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.  
 CC The invention also includes fragments and biologically functional  
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are  
 CC useful for determining the bioavailability of drugs and for  
 CC screening PGP inhibitors. They are useful for the diagnosis and  
 CC treatment of conditions characterised by PGP activity, by  
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids  
 CC are used as oligonucleotide probes. Complements of PGP nucleic  
 CC acids are useful as antisense oligonucleotides, to induce a PGP  
 CC "knockout" phenotype. They are used to prepare a non-human  
 CC transgenic animal, which are valuable as genetic models for  
 CC human diseases.  
 CC The present sequence is dog P-glycoprotein (PGP) allelic variant.  
 CC This sequence is also referred as Genotype D protein. The PGP  
 CC enzyme functions as an efflux pump exporting small molecules  
 CC across the cell membrane. This enzyme is a member of the ABC

CC transporter family.  
 XX  
 SQ Sequence 1281 AA;  
 Query Match 12.5%; Score 69; DB 22; Length 1281;  
 Best Local Similarity 25.5%; Pred. No. 70;  
 Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFFSVITVMTLIMASEMV-----NSAFTVWSGPGCNRARERYSKCGSAIHOKGYD 59  
 DB 336 VFESVLIGAFSIGQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGHKKPDNIKMLE 395  
 QY 60 -----FSY-----TGQTALYNQAGC 75  
 DB 396 FKXVHFSPSRKEVKILKGLNLKQSGQTVALVGNSSG 433

Search completed: January 12, 2003, 09:35:53  
 Job time : 68 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:36:45 (Search time 25 Seconds)

(without alignments)  
120.046 Million cell updates/sec

Title: US-09-882-434a-1

Perfect score: 551  
Sequence: 1 MASTKLFVETVITMMLIAMA.....FGSSARACNPFQKSIQIC 102

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/PCCTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74.5	13.5	486	1	US-07-872-678A-48
2	68	12.3	1280	2	US-08-583-276-19
3	68	12.3	1280	2	US-08-752-447-2
4	68	12.3	1280	4	US-09-316-167-2
5	68	12.3	1280	6	5206352-4
6	67.5	12.3	902	1	US-08-701-846-2
7	63.5	11.5	1019	1	US-08-296-014A-4
8	63.5	11.5	1019	2	US-08-596-405-4
9	63.5	11.5	1019	2	US-08-877-620-4
10	63.5	11.5	1083	1	US-08-296-014A-2
11	63.5	11.5	1083	2	US-08-596-405-2
12	63.5	11.5	1083	2	US-08-877-620-2
13	62.5	11.3	135	2	US-07-857-224B-99
14	62	11.3	442	4	US-09-347-833-11
15	62	11.3	3729	2	US-08-804-227C-4
16	61.5	11.2	111	2	US-07-857-224B-105
17	61.5	11.2	115	2	US-07-857-224B-97
18	61.5	11.2	135	2	US-07-857-224B-98
19	61	11.1	510	4	US-09-212-247C-5
20	61	11.1	1196	1	US-08-144-121-4
21	61	11.1	1196	2	US-08-735-893-4
22	61	11.1	1279	2	US-08-784-649A-2
23	60.5	11.0	985	2	US-08-880-326-41
24	60.5	11.0	1349	2	US-08-612-734B-2
25	60	10.9	687	1	US-08-470-702-10
26	60	10.9	687	1	US-08-467-831-10
27	60	10.9	730	2	US-08-696-944-2

28	60	10.9	871	4	US-09-245-041-19	Sequence 19, Appl
29	60	10.9	1198	4	US-09-245-041-131	Sequence 131, App
30	60	10.9	1198	4	US-09-794-236-3	Sequence 3, Appl1
31	60	10.9	1350	4	US-09-245-041-17	Sequence 17, Appl
32	60	10.9	2787	4	US-09-245-041-15	Sequence 15, Appl
33	59.5	10.8	305	4	US-09-230-232-1	Sequence 1, Appl1
34	59.5	10.8	861	4	US-08-960-048-12	Sequence 12, Appl
35	59	10.7	352	2	US-08-483-926A-11	Sequence 11, Appl
36	59	10.7	549	4	US-09-245-041-9	Sequence 9, Appl1
37	59	10.7	687	1	US-08-204-656B-10	Sequence 10, Appl
38	59	10.7	1260	4	US-09-245-041-2	Sequence 2, Appl1
39	59	10.7	1345	2	US-08-977-767-3	Sequence 3, Appl1
40	58.5	10.6	415	4	US-09-134-001C-3957	Sequence 3957, Ap
41	58.5	10.6	510	4	US-09-134-001C-4541	Sequence 4541, Ap
42	58.5	10.6	585	4	US-09-620-412C-337	Sequence 337, App
43	58.5	10.6	1404	2	US-08-400-159-2	Sequence 2, Appl1
44	58.5	10.6	1404	3	US-08-611-729A-2	Sequence 2, Appl1
45	58.5	10.6	1752	4	US-09-556-877-180	Sequence 180, App

#### ALIGNMENTS

RESULT 1  
US-07-872-678A-48  
Sequence 48, Application US/07872678A  
Patent No. 5541060  
GENERAL INFORMATION:  
APPLICANT: Bell, Graeme, et al.  
TITLE OF INVENTION: DETECTION OF EARLY-ONSET  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Arnold, White & Durkee  
STREET: Post Office Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/872,678A  
FILING DATE: 22-APRIL-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coughlin, Daniel F.  
REGISTRATION NUMBER: 36,111  
REFERENCE/DOCKET NUMBER: ARCD016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-787-1400  
TELEFAX: 713-789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-872-678A-48

Query Match 13.5%, Score 74.5, DB 1, Length 486,  
Best Local Similarity 39.3%, Pred. No. 1.1,  
Matches 24, Conservative 4, Mismatches 18, Indels 15, Gaps 4;  
OY 40 RAERYKSGCSATHHKGVDYSGTAAALYNQAGCGVAHTRF-GSARACNP-----FG 94  
DB 391 RAARSTVSGIAHICQKRGK--TGHIAA-----DGSVSTRYPGRKRAALAKOIVG 440

```
QY 95 W 95
Db 441 W 441

RESULT 2
US-08-583-276-19
; Sequence 19, Application US/08583276
; Patent No. 5837536
; GENERAL INFORMATION:
; APPLICANT: McDonagh, Kevin T.
; APPLICANT: Nienhuis, Arthur
; APPLICANT: Tolstoshev, Paul
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
; TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
; TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi & Stewart
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: DM4 V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,276
; FILING DATE: 05-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/332,444
; FILING DATE: 31-OCT-1994
; APPLICATION NUMBER: 07/887,712
; FILING DATE: 22-MAY-1992
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; US-08-583-276-19

Query Match 12.3%; Score 68; DB 2; Length 1280;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFESVITVMILIAMASEMV-----NGSAFTVWSGPGCNRAERYSKGCSAIHOKGYD 59
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 VFFSVLIGAFSGVQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGKHGKPDNIKGNLE 393
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 -----FSY-----TGQTAALYNQAGC 75
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 FRNVHFSYPSRKEVKILKGLNLKQSGQTVALVGNSGC 431

RESULT 3
US-08-752-447-2
; Sequence 2, Application US/08752447
; Patent No. 5994088
; GENERAL INFORMATION:
; APPLICANT: Mechetner, Eugene
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods and Reagents for Preparing and
; TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
; STREET: 300 South Wacker Drive, Seventh Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/316.167
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/752,447
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; US-08-752-447-2
```

```
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
; STREET: 300 South Wacker Drive, Seventh Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,447
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5994088nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 95,1121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-752-447-2

Query Match 12.3%; Score 68; DB 2; Length 1280;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFESVITVMILIAMASEMV-----NGSAFTVWSGPGCNRAERYSKGCSAIHOKGYD 59
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 VFFSVLIGAFSGVQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGKHGKPDNIKGNLE 393
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 -----FSY-----TGQTAALYNQAGC 75
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 FRNVHFSYPSRKEVKILKGLNLKQSGQTVALVGNSGC 431

RESULT 4
US-09-316-167-2
; Sequence 2, Application US/09316167
; Patent No. 6365357
; GENERAL INFORMATION:
; APPLICANT: Mechetner, Eugene
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods and Reagents for Preparing and
; TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
; STREET: 300 South Wacker Drive, Seventh Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/316.167
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/752,447
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
```

NAME: No. 6365357nan, Kevin E  
 REGISTRATION NUMBER: 35,303  
 REFERENCE/DOCKET NUMBER: 95,1121  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312-913-0001  
 TELEFAX: 312-913-9808  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1280 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-316-167-2

Query Match 12.3%; Score 68; DB 4; Length 1280;  
 Best Local Similarity 24.5%; Pred. No. 21;  
 Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LPEFVITVMMLIAMAEMV-----NGSAFTVWSGPGCNRRAERYSKGCGSAIHQKGYD 59  
 DB 334 VFESVLIGAFSGVQASPSIEAFANARGAAYEIFKIIDNKPSIDYSKSGHKPDNIKGNLE 393

QY 60 -----FSY-----TGOTATALYNQAGC 75  
 DB 394 FRNVHFSYPSRKEVKILKGLNLKYGSGQTVALVGNSSC 431

RESULT 5  
 5206352-4  
 Patent No. 5206352  
 APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman, Michael M.

TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA  
 SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS  
 NUMBER OF SEQUENCES: 4  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/622,836  
 FILING DATE: 24-SEP-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 892,575  
 FILING DATE: 01-AUG-1986  
 APPLICATION NUMBER: 845,610  
 FILING DATE: 28-MAR-1986  
 SEQ ID NO: 4:  
 LENGTH: 1280

5206352-4

Query Match 12.3%; Score 68; DB 6; Length 1280;  
 Best Local Similarity 24.5%; Pred. No. 21;  
 Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LPEFVITVMMLIAMAEMV-----NGSAFTVWSGPGCNRRAERYSKGCGSAIHQKGYD 59  
 DB 334 VFESVLIGAFSGVQASPSIEAFANARGAAYEIFKIIDNKPSIDYSKSGHKPDNIKGNLE 393

QY 60 -----FSY-----TGOTATALYNQAGC 75  
 DB 394 FRNVHFSYPSRKEVKILKGLNLKYGSGQTVALVGNSSC 431

RESULT 6  
 US-08-701-846-2  
 Sequence 2, Application US/08701846  
 Patent No. 571069  
 GENERAL INFORMATION:

APPLICANT: Granados, Robert R.  
 TITLE OF INVENTION: DNA SEQUENCE CODING FOR A POLYPEPTIDE  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Barnard, Brown & Michaelis  
 STREET: 306 E. State St., Suite 220  
 CITY: Ithaca,

STATE: NY.  
 COUNTRY: USA  
 ZIP: 14850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/701,846  
 FILING DATE: 23-AUG-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/002,743  
 FILING DATE: 24-AUG-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michaels, Christopher A.  
 REGISTRATION NUMBER: 34,390  
 REFERENCE/DOCKET NUMBER: BTI-32  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (607)273-1711  
 TELEFAX: (607)273-2609  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 902 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-701-846-2

Query Match 12.3%; Score 67.5; DB 1; Length 902;  
 Best Local Similarity 28.6%; Pred. No. 16;  
 Matches 20; Conservative 10; Mismatches 33; Indels 7; Gaps 4;

QY 26 GSAP--TWSGPGCNRRAE--RYSKCGCGSAIHQK-GYFSTGTGOT--AALYNQAGCSGV 78  
 DB 213 GGAIVYGAFWTAPASTVLGELRVSPPTMMVVIHELGHAYDFVTVTRLIEIWNNSPCDRI 272

QY 79 AATFGSSAR 88  
 DB 273 QYTWMNKTKR 282

RESULT 7  
 US-08-296-014A-4  
 Sequence 4, Application US/08296014A  
 Patent No. 5716834  
 GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling  
 APPLICANT: Ho, Bow  
 TITLE OF INVENTION: The Cloned Factor C cDNA of the  
 TITLE OF INVENTION: Singapore Horseshoe Crab, *Caridioscorpius*  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Birch, Stewart, Kolasch & Birch  
 STREET: 8110 Gatehouse Road, Suite 500 East  
 CITY: Falls Church  
 STATE: Virginia  
 COUNTRY: USA  
 ZIP: 22042

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/296,014A  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy, Jr., Gerald M.  
 REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1019 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-296-014A-4

Query Match 11.5%; Score 63.5; DB 1; Length 1019;

Best Local Similarity 32.8%; Pred. No. 54;

Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

Qy 32 WSG--PGCNNRERYSKGCSATHQKGYDF-SYTGQTALYNQAGCGVAHTRFGSS 86

Db 314 WSGQSPSCVKVADREVDKSKAV-----DFLDDVGEPVRIHCPAGCSLTAGTVMGTA 365

RESULT 8

US-08-596-405-4

Sequence 4, Application US/08596405

Patent No. 5858706

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling

TITLE OF INVENTION: The Cloned Factor C CDNA of the

TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus

TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch &amp; Birch

STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/596,405

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1019 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-596-405-4

Query Match 11.5%; Score 63.5; DB 2; Length 1019;

Best Local Similarity 32.8%; Pred. No. 54;

Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

Qy 32 WSG--PGCNNRERYSKGCSATHQKGYDF-SYTGQTALYNQAGCGVAHTRFGSS 86

Db 314 WSGQSPSCVKVADREVDKSKAV-----DFLDDVGEPVRIHCPAGCSLTAGTVMGTA 365

RESULT 9

US-08-877-620-4

Sequence 4, Application US/08877620

Patent No. 5985590

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling

APPLICANT: Ho, Bow

TITLE OF INVENTION: The Cloned Factor C CDNA of the

TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus

TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch &amp; Birch

STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/877,620

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/596,405

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1019 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-877-620-4

Query Match 11.5%; Score 63.5; DB 2; Length 1019;

Best Local Similarity 32.8%; Pred. No. 54;

Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

Qy 32 WSG--PGCNNRERYSKGCSATHQKGYDF-SYTGQTALYNQAGCGVAHTRFGSS 86

Db 314 WSGQSPSCVKVADREVDKSKAV-----DFLDDVGEPVRIHCPAGCSLTAGTVMGTA 365

RESULT 10

US-08-296-014A-2

Sequence 2, Application US/08296014A

Patent No. 5716834

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling

APPLICANT: Ho, Bow

TITLE OF INVENTION: The Cloned Factor C CDNA of the

TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus

TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch &amp; Birch

STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: USA



ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,014A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy, Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1083 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-296-014A-2

Query Match 11.5%; Score 63.5; DB 1; Length 1083;  
Best Local Similarity 32.8%; Pred. No. 58;  
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

QY 32 MSG--PGCNRAERYSKGCSAIHQKGYDF-STYGQTAALYNQAGCGVAHTRFGSS 86  
DB 378 MSGQSPSCVKVADREVDCSKAV-----DFLDVGEVRIHCPAGCSLTGTWGTGA 429

RESULT 11  
US-08-596-405-2  
Sequence 2, Application US/08596405  
Patent No. 5858706  
GENERAL INFORMATION:  
APPLICANT: Ding, Jeak Ling  
TITLE OF INVENTION: The Cloned Factor C cDNA of the  
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus  
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,405  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy, Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1083 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-596-405-2

Query Match 11.5%; Score 63.5; DB 2; Length 1083;  
Best Local Similarity 32.8%; Pred. No. 58;  
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;  
QY 32 MSG--PGCNRAERYSKGCSAIHQKGYDF-STYGQTAALYNQAGCGVAHTRFGSS 86  
DB 378 MSGQSPSCVKVADREVDCSKAV-----DFLDVGEVRIHCPAGCSLTGTWGTGA 429

RESULT 12  
US-08-877-620-2  
Sequence 2, Application US/08877620  
Patent No. 5985590  
GENERAL INFORMATION:  
APPLICANT: Ding, Jeak Ling  
TITLE OF INVENTION: The Cloned Factor C cDNA of the  
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus  
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/877,620  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/596,405  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy, Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1083 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-877-620-2

Query Match 11.5%; Score 63.5; DB 2; Length 1083;  
Best Local Similarity 32.8%; Pred. No. 58;  
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;  
QY 32 MSG--PGCNRAERYSKGCSAIHQKGYDF-STYGQTAALYNQAGCGVAHTRFGSS 86  
DB 378 MSGQSPSCVKVADREVDCSKAV-----DFLDVGEVRIHCPAGCSLTGTWGTGA 429

RESULT 13  
US-07-857-224B-99  
Sequence 99, Application US/07857224B  
Patent No. 5958784

```

; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Odell, Joan T.
; TITLE OF INVENTION: Factors Involved in Gene Expression
; FILE REFERENCE: BB-1172
; CURRENT APPLICATION NUMBER: US/09/347,833

```

Qy	24	VNGSAFTVMSG-PCGCNNRAEY\$KCGCSAIHQKGGYDFS YTGQTAAALYNQAGCSVAH--	80
Db	1343	VTSVAWTPWLGTPAADSLGEOM\$RAGITPLDPAASLD-----ALARAVGRRAGCTVTVADID	13

Mon Jan 13 09:25:42 2003

us-09-882-434a-1.rat

Page 7

OY 81 -TRFGSSARACNP 92  
|||:  
Db 1399 WERFASAYTATRP 1411

Search completed: January 12, 2003, 09:40:11  
Job time : 28 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:39:30 ; Search time 20 Seconds

(without alignments)  
98.945 Million cell updates/sec

Title: US-09-882-434A-1

Sequence: 1 MASTKLFPSVITVMTLIMAMA.....FGSSARACNPFKWSIFIOC 102

Scoring table: BIOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551	100.0	102	US-09-882-434A-1	Sequence 1, Appli
2	426	77.3	76	US-09-882-434A-15	Sequence 15, Appli
3	426	77.3	76	US-09-882-434A-16	Sequence 16, Appli
4	426	77.3	76	US-09-882-434A-17	Sequence 17, Appli
5	421	76.0	76	US-09-882-434A-19	Sequence 19, Appli
6	419	76.0	76	US-09-882-434A-18	Sequence 18, Appli
7	417	75.7	76	US-09-882-434A-21	Sequence 21, Appli
8	415	75.3	76	US-09-882-434A-20	Sequence 20, Appli
9	75.5	13.7	486	US-09-801-368-154	Sequence 154, App
10	70.5	12.8	267	US-09-764-868-1164	Sequence 1164, Ap
11	69	12.5	9	US-10-044-671-2	Sequence 2, Appli
12	68	12.3	1280	US-10-072-621-7	Sequence 7, Appli
13	68	12.3	1280	US-09-866-866A-2	Sequence 2, Appli
14	68	12.3	1280	US-09-866-866A-4	Sequence 4, Appli
15	64.5	11.7	9	US-10-016-634A-115	Sequence 115, App
16	64	11.6	1272	US-09-769-097-2	Sequence 2, Appli
17	64	11.6	1272	US-09-769-097-4	Sequence 4, Appli
18	64	11.6	1609	US-09-938-275-11	Sequence 11, Appli
19	61	11.1	501	US-08-635-967-2	Sequence 2, Appli

20	61	11.1	914	9	US-09-975-143-47	Sequence 47, Appli
21	61	11.1	1786	10	US-09-873-676-113	Sequence 113, App
22	61	11.1	1786	10	US-09-938-275-6	Sequence 6, Appli
23	60.5	11.0	94	9	US-09-950-933A-44	Sequence 44, Appli
24	60.5	11.0	348	10	US-09-791-961-2	Sequence 2, Appli
25	60.5	11.0	787	9	US-09-712-363-178	Sequence 178, App
26	60	10.9	515	10	US-09-895-072-6	Sequence 6, Appli
27	60	10.9	515	10	US-09-986-552-6	Sequence 6, Appli
28	60	10.9	871	10	US-09-893-238-19	Sequence 19, Appli
29	60	10.9	1276	10	US-09-866-866A-8	Sequence 8, Appli
30	60	10.9	1350	10	US-09-893-238-17	Sequence 17, Appli
31	60	10.9	1607	10	US-09-938-275-10	Sequence 10, Appli
32	60	10.9	1786	10	US-09-938-275-7	Sequence 7, Appli
33	60	10.9	2787	10	US-09-893-238-15	Sequence 15, Appli
34	59.5	10.8	861	10	US-09-838-539-12	Sequence 12, Appli
35	59.5	10.8	1404	10	US-09-862-027-24	Sequence 24, Appli
36	59.5	10.8	1743	12	US-10-052-586-451	Sequence 451, App
37	59	10.7	549	10	US-09-893-238-9	Sequence 9, Appli
38	59	10.7	1260	10	US-09-893-238-2	Sequence 2, Appli
39	58.5	10.6	476	12	US-10-078-929-201	Sequence 201, App
40	58.5	10.6	585	10	US-09-841-132-337	Sequence 337, App
41	58.5	10.6	1404	10	US-09-944-849-8	Sequence 8, Appli
42	58.5	10.6	1431	9	US-09-843-930A-2	Sequence 2, Appli
43	58.5	10.6	1752	10	US-09-841-132-180	Sequence 180, App
44	58.5	10.6	5405	9	US-10-023-380-1116	Sequence 1116, Ap
45	58.5	10.6	5405	10	US-09-922-217-1116	Sequence 1116, Ap

## ALIGNMENTS

```
RESULT 1
US-09-882-434A-1
; Sequence 1, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie Lynn
; TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
; FILE REFERENCE: CULIN8.1CPICI
; CURRENT APPLICATION NUMBER: US/09/882,434A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/364395
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/117615
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Macadamia integrifolia
US-09-882-434A-1

Query Match      100.0%; Score 551; DB 10; Length 102;
Best Local Similarity 100.0%; Pred. No. 4,1e-57;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTKLFPSVITVMTLIMASEMNGSAFTWISGRCNNRERISKCCGSAIHQKGYDF 60
DB 1 MASTKLFPSVITVMTLIMASEMNGSAFTWISGRCNNRERISKCCGSAIHQKGYDF 60
QY 61 SYTGOTALYNQAGCGGVAHTRFSSARACNPFKWSIFIOC 102
DB 61 SYTGOTALYNQAGCGGVAHTRFSSARACNPFKWSIFIOC 102
```

Mi39K

APPLICANT: Green, Jodie Lyn  
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN  
FILE REFERENCE: CULN18.1CPI1  
CURRENT APPLICATION NUMBER: US/09/882,434A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/364395  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/117615  
PRIOR FILING DATE: 1998-11-09  
PRIOR APPLICATION NUMBER: PCT/AU97/00052  
PRIOR FILING DATE: 1997-01-31  
PRIOR APPLICATION NUMBER: AU PN 7802  
PRIOR FILING DATE: 1996-01-31  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: M154K variant. Variant M154K  
OTHER INFORMATION: containing a Lysine at amino acid 54 (used primer)  
OTHER INFORMATION: from SEQ ID NO:12 to produce).  
US-09-882-434A-19

Query Match 76.4%; Score 421; DB 10; Length 76;  
Best Local Similarity 98.7%; Pred. No. 3.5e-42;  
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 SAFTWVGPGCNRNRAERYSKCGCSAIIHQKGYDPSYTGQTALVYNQAGCGVAATRRGSS 86  
DB 1 SAFTWVGPGCNRNRAERYSKCGCSAIIHQKGYDPSYTGQTALVYNQAGCGVAATRRGSS 60  
QY 87 ARACNPFGMKSIPIQC 102  
DB 61 ARACNPFGMKSIPIQC 76

RESULT 6  
US-09-882-434A-18  
Sequence 18, Application US/09882434A  
Patent No. US20020108144A1  
GENERAL INFORMATION:  
APPLICANT: Manners, John M.  
APPLICANT: Marcus, John Paul  
APPLICANT: Goulter, Kenneth C.  
APPLICANT: Green, Jodie Lyn  
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN  
FILE REFERENCE: CULN18.1CPI1  
CURRENT APPLICATION NUMBER: US/09/882,434A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/364395  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/117615  
PRIOR FILING DATE: 1998-11-09  
PRIOR APPLICATION NUMBER: PCT/AU97/00052  
PRIOR FILING DATE: 1997-01-31  
PRIOR APPLICATION NUMBER: AU PN 7802  
PRIOR FILING DATE: 1996-01-31  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: M154V variant. Variant M154V  
OTHER INFORMATION: containing a Valine at amino acid 54 (used primer)  
OTHER INFORMATION: from SEQ ID NO:11 to produce).  
US-09-882-434A-18

Query Match 76.0%; Score 419; DB 10; Length 76;  
Best Local Similarity 98.7%; Pred. No. 6e-42;

Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 27 SAFTWVGPGCNRNRAERYSKCGCSAIIHQKGYDPSYTGQTALVYNQAGCGVAATRRGSS 86  
DB 1 SAFTWVGPGCNRNRAERYSKCGCSAIIHQKGYDPSYTGQTALVYNQAGCGVAATRRGSS 60  
QY 87 ARACNPFGMKSIPIQC 102  
DB 61 ARACNPFGMKSIPIQC 76

RESULT 7  
US-09-882-434A-21  
Sequence 21, Application US/09882434A  
Patent No. US20020108144A1  
GENERAL INFORMATION:  
APPLICANT: Manners, John M.  
APPLICANT: Marcus, John Paul  
APPLICANT: Goulter, Kenneth C.  
APPLICANT: Green, Jodie Lyn  
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN  
FILE REFERENCE: CULN18.1CPI1  
CURRENT APPLICATION NUMBER: US/09/882,434A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/364395  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/117615  
PRIOR FILING DATE: 1998-11-09  
PRIOR APPLICATION NUMBER: PCT/AU97/00052  
PRIOR FILING DATE: 1997-01-31  
PRIOR APPLICATION NUMBER: AU PN 7802  
PRIOR FILING DATE: 1996-01-31  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 21  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: M146K/54K variant. Variant M146K/54K  
OTHER INFORMATION: containing a Lysine at amino acid 46 and  
OTHER INFORMATION: a Lysine at amino acid 54.  
US-09-882-434A-21

Query Match 75.7%; Score 417; DB 10; Length 76;  
Best Local Similarity 97.4%; Pred. No. 1e-41;  
Matches 74; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 27 SAFTWVGPGCNRNRAERYSKCGCSAIIHQKGYDPSYTGQTALVYNQAGCGVAATRRGSS 86  
DB 1 SAFTWVGPGCNRNRAERYSKCGCSAIIHQKGYDPSYTGQTALVYNQAGCGVAATRRGSS 60  
QY 87 ARACNPFGMKSIPIQC 102  
DB 61 ARACNPFGMKSIPIQC 76

RESULT 8  
US-09-882-434A-20  
Sequence 20, Application US/09882434A  
Patent No. US20020108144A1  
GENERAL INFORMATION:  
APPLICANT: Manners, John M.  
APPLICANT: Marcus, John Paul  
APPLICANT: Goulter, Kenneth C.  
APPLICANT: Green, Jodie Lyn  
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN  
FILE REFERENCE: CULN18.1CPI1  
CURRENT APPLICATION NUMBER: US/09/882,434A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/364395  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/117615

```
Query Match      12.5%; Score 69; DB 9; Length 1280;
Best Local Similarity    25.5%; Prd. NO. 6;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;
```

QY        6   LFPSVITVMILIAMASEMV-----NGSFTVMVGPGCCNNRAERYSKCSCCSAIHOKGYD 59  
           | ||| : | : : : : | : : : :  
Db      335 VFPSVLIGAFSIGOASPISIEAFANARAAEYIKPIDNKPIDSYSKSHKPDNIKGNLE 394

QY      60 -----FSY-----TCTAALYNAGC 75





Db 65 GSSAWPDSTCN 75

Search completed: January 12, 2003, 09:46:46  
Job time : 22 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:36:00 ; Search time 43 seconds  
(without alignments)  
228.040 Million cell updates/sec

Title: US-09-882-434a-1

Perfect score: 551  
Sequence: 1 MASTKLFSSVITVMTLIMA.....FGSSARACNPFQWKSIFIOC 102

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR\_73.\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*  
5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.5	13.7	315	2 T35804	probable aldehyde
2	75.5	13.7	486	2 KIBYHB	hexokinase (EC 2.7
3	69.5	12.6	532	2 AB3552	alcohol dehydrogen
4	69.5	12.5	1281	2 T48123	p-glycoprotein iso
5	68.5	12.4	1557	2 T28811	hypothetical prote
6	68.5	12.3	1280	1 DVHUI	multidrug resistan
7	67.5	12.3	184	2 T16044	hypothetical prote
8	66.5	12.1	81	2 T02667	proteinase inhibit
9	66.5	12.1	151	2 JCG040	fibrin protein ag
10	66.5	12.1	157	2 S31078	seed allergen RA5
11	66.5	12.1	162	2 S21157	seed allergen RA17
12	66.5	12.1	251	2 T34656	hypothetical prote
13	66.5	12.1	461	2 AE0311	MADH2 dehydrogenas
14	66.5	12.0	1276	1 DVMS2	multidrug resistan
15	65.5	11.9	151	2 AH0635	nucleation compone
16	65.5	11.9	233	2 T23296	hypothetical prote
17	65.5	11.9	244	2 T04212	osmotic precursor
18	65.5	11.8	165	2 S31080	seed allergen RA14
19	64.5	11.7	485	1 KIBYHA	hexokinase (EC 2.7
20	64.5	11.6	320	2 S00935	ribose protein - Stap
21	64.5	11.6	1378	2 S41646	p-glycoprotein - r
22	64.5	11.6	1551	2 AB2410	WD-repeat protein
23	64.5	11.6	1609	1 KIBYHB	laminin gamma-1 ch
24	64.5	11.6	3016	2 S77300	hypothetical prote
25	63.5	11.5	166	2 S31082	seed allergen RA62
26	63.5	11.5	503	2 S31940	starch-degrading e
27	63.5	11.5	503	2 C97262	N-terminal domain
28	63.5	11.5	828	1 D31442	outer membrane ush
29	63.5	11.5	1019	2 A38738	coagulation factor

#### ALIGNMENTS

30	63	11.4	157	2 T02664	allergen - rice
31	63	11.4	159	2 S26238	pathogenesis-relat
32	63	11.4	339	2 A47611	env polyprotein -
33	63	11.4	1276	1 DVHUI	multidrug resistan
34	63	11.4	1279	1 DVHUI	multidrug resistan
35	62.5	11.3	125	2 T16247	hypothetical prote
36	62.5	11.3	166	2 S59922	allergen RA14B pre
37	62.5	11.3	168	2 T07146	pathogenesis-relat
38	62.5	11.3	235	1 OTTC2	chaumatin II precu
39	62.5	11.3	779	2 H71301	probable membrane
40	62	11.3	159	2 VCT014	pathogenesis-relat
41	62	11.3	159	2 S00900	heat shock 18k pro
42	62	11.3	234	2 T21495	hypothetical prote
43	62	11.3	340	2 T13781	MADH2 dehydrogenas
44	62	11.3	461	2 C86935	probable carboxype
45	62	11.3	553	2 T52362	hypothetical prote

#### RESULT 1

T35804  
probable aldehyde dehydrogenase - Streptomyces coelicolor (fragment)

C/Species: Streptomyces coelicolor  
C/Date: 05-Nov-1999 #sequence revision 05-Nov-1999 #text\_change 05-Nov-1999

C/Accession: T35804

R/Murphy, L., Harris, D., Parkhill, J., Barrell, B.G., Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998

A/Reference number: Z21556

A/Accession: T35804

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-315 <MUR>

A/Cross-references: EMBL:AL031225; PIDN:CAA20223.1; GSPDB:GN00070; SCODEB:SC8B7.12c

A/Experimental source: strain A3(12)

C/Genetics:

A/Gene: SCODEB:SC8B7.12c

Query Match

Best Local Similarity 13.7%; Score 75.5; DB 2; Length 315;

Matches 30; Conservative 20; Mismatches 47; Indels 19; Gaps 5;

Db

1 MASTKLFSSVITVMTL-----TAMASEMNGSAFTVWGPQCN-NRARIYKSGCSATH 53

Db

199 VVSEIRFGVPLVLPDPTDDEGIRLNDPRYGLAASMSRDVRRARATREIRAGCWAIN 258

Qy

54 Q-----KGGVDFSYTGOTATLVNAGCSGVAHTRFGSSARACNPFQW-KSIF 99

Db

259 DHPIISEMHPHGYKASGFGKMSAISFEYTYQKVMFNTVAAK--DMHRTVF 312

RESULT 2

KIBYHB

hexokinase (EC 2.7.1.1) B - yeast (Saccharomyces cerevisiae)

N/Alternate names: HEX1 protein; hexokinase II; hexokinase PII; protein G0556; protein N

C/Species: Saccharomyces cerevisiae

C/Date: 28-Dec-1987 #sequence revision 12-Apr-1996 #text\_change 16-Jun-2000

C/Accession: S61608; B23523; S22430; A23958; S05731; S33656; S64279; A53632; S28555

R/Colisac: E; Maillet, E.; Robineau, S.; Netter, P.

submitted to the EMBL Data Library, December 1995

A/Reference number: S61598

A/Accession: S61608

A/Molecule type: DNA

A/Residues: 1-486 <COI>

A/Cross-references: EMBL:X04357; NID:G1150575; PIDN:CAA64134.1; PID:G1150586

R/Stachel, C.; Stachel, J.; Swan, J.; Botstein, D.; Konigsberg, W.

Nucleic Acids Res. 14, 945-963, 1986

A/Title: Identification, cloning and sequence determination of the genes specifying hexo

A/Reference number: A93649; MUID:86120382; PMID:3003701

A/Accession: B23523

A/Molecule type: DNA

A/Residues: 1-28, 'I', '30-60, 'V', '62-196, 'S', '198-486 <STA>

A:Cross-references: EMBL:X03483; NID:g3792; PIDN:CAA27203.1; PID:g3793  
A:Note: the authors translated the codon GTT for residue 61 as Gly  
A:Accession: S22430  
A:Molecule type: protein  
A:Residues: 2-7; I: 30-46; 78-111; 114-119; 166-173; 177-194; 228-236; 250-281; 283-300; 330-335  
R:Prohlich, K.; Entian, K.; Mecke, D.  
Gene 36, 105-111, 1985  
A:Title: The primary structure of the yeast hexokinase PII gene (HXK2) which is responsible for the regulation of glucose uptake in *Saccharomyces cerevisiae*  
A:Reference number: A23958; MUID:186056943; PMID:3905511  
A:Accession: A23958  
A:Molecule type: DNA  
A:Residues: 1-32; N: 34-60; V: 62-420; ST: 423-443; PH: 446-452; V: 454-461; P: 463-486  
A:Cross-references: EMBL:M11181  
A:Note: the authors translated the codon GTT for residue 61 as Gly  
R:Schmidt, J.J.; Colowick, S.P.  
Arch. Biochem. Biophys. 158, 471-477, 1973  
A:Title: Identification of a peptide sequence involved in association of subunits of yeast hexokinase  
A:Reference number: S05731; MUID:74114889; PMID:4592981  
A:Accession: S05731  
A:Molecule type: protein  
A:Residues: 2-12 <SCH>  
R:Bretwieser, W.; Price, C.; Schuster, T.  
Yeast 9, 551-556, 1993  
A:Title: Identification of a gene encoding a novel zinc finger protein in *Saccharomyces cerevisiae*  
A:Reference number: S33654; MUID:93311123; PMID:8322518  
A:Accession: S33654  
A:Molecule type: DNA  
A:Residues: 1-247 <BRE>  
A:Cross-references: EMBL:X67787; NID:g3707; PIDN:CAA48003.1; PID:g3710  
R:Coissac, E.; Maillier, E.; Netter, P.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64271  
A:Accession: S64271  
A:Molecule type: DNA  
A:Residues: 1-486 <COW>  
A:Cross-references: EMBL:272775; NID:gl3222930; PIDN:CAA96973.1; PID:gl3222931; GSPDB:GN00190  
A:Experimental source: strain S288C  
R:Kriegel, T.M.; Rush, J.; Vojtek, A.B.; Clifton, D.; Fraenkel, D.G.  
Biochemistry 33, 148-152, 1994  
A:Title: In vivo phosphorylation site of hexokinase 2 in *Saccharomyces cerevisiae*  
A:Reference number: A53632; MUID:94114477; PMID:8286332  
A:Accession: A53632  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 13-21 <KRI>  
C:Genetics:  
A:Gene: SGD.HXK2; HEX1; SC12; HXB; MIPS:YGL253W  
A:Cross-references: SGD:S0003222; MIPS:YGL253W  
A:Map position: 7L  
C:Superfamily: hexokinase; hexokinase homology  
C:Keywords: allosteric regulation; ATP; glycolysis; phosphoprotein; phosphotransferase  
F:2-486/Product: hexokinase B #status experimental <MAT>  
F:36-470/Domain: hexokinase homology <HXK>  
F:15/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 13.7%; Score 75.5; DB 1; Length 486;  
Best Local Similarity 30.2%; Pred. No. 3.5;  
Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;  
QY 8 FSVITVMM---LIAMSEWNGSAFTVWSGPGCNNRAERYSKGCSAIHQKGYDFSHTG 64  
DB 369 FGINTVQERKLIRLSLIGA-----RAARLSVCGIATCQKRGYK---TG 412  
QY 65 QTAALYNQAGCGVAHTRF-GSSARACNP---FGW 95  
DB 413 HIAA-----DGSVYNYRPFKKAALAKDIYGV 441

RESULT 3  
AB3552  
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: A23252; PMID:11756688  
A:Accession: AB3552  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-532 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL53581.1; PID:gl7984492; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10339  
A:Map position: II  
C:Superfamily: alcohol oxidase  
C:Keywords: oxidoreductase

Query Match 12.6%; Score 69.5; DB 2; Length 532;  
Best Local Similarity 27.2%; Pred. No. 16;  
Matches 22; Conservative 8; Mismatches 16; Indels 35; Gaps 4;  
QY 29 FTVWSGPGCNNRAERYSK---CGCSAIH-----OKGYDFSHTGTAALYNQAGCG 77  
DB 64 FTTEAEPLNGRSLNTPRGKVLGGCSINGMIYMRGQACDYD-----LWRQAGCDG 114  
QY 78 VAHTRFGSSARACNPFQWKSII 98  
DB 115 -----NGWDDVV 120

RESULT 4  
I48123  
p-glycoprotein isoform III - Chinese hamster  
C:Species: *Cricetulus griseus* (Chinese hamster)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 02-Feb-2001  
C:Accession: I48123  
R:Endicott, J.A.; Sarangi, P.; Ling, V.  
DNA Seq. 2, 89-101, 1991  
A:Title: Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene family  
A:Reference number: I48121; MUID:92135896; PMID:1685679  
A:Accession: I48123  
A>Status: preliminary; translated from GB/EMBL/DBS  
A:Molecule type: mRNA  
A:Residues: 1-1281 <RES>  
A:Cross-references: GB:M60042; NID:gl91168; PIDN:AAA68885.1; PID:gl91169  
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology  
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop  
F:412-606/Domain: ATP-binding cassette homology <ABC1>  
F:429-436/Region: nucleotide-binding motif A (P-loop)  
F:1054-1250/Domain: ATP-binding cassette homology <ABC2>  
F:1071-1078/Region: nucleotide-binding motif A (P-loop)

Query Match 12.5%; Score 69; DB 2; Length 1281;  
Best Local Similarity 21.4%; Pred. No. 41;  
Matches 21; Conservative 17; Mismatches 32; Indels 28; Gaps 2;  
QY 6 LFPSVITVMMLIAMSEWNV-----GSAFTVWSGPGCNNRAERYSKGCSAIHQKGYD 59  
DB 336 VFPSILIGAFVGOAPCIDAFANARGAAYVIFDIIDNNPKIDSFSERGHKPDISKGNLD 395  
QY 60 FS-----YTGOTAAALYNQAGC 75  
DB 396 FSDVHFSYPSRANIKLGLNLKVSQGTVALVGNSSG 433

RESULT 5  
T28811  
hypothetical protein C54D1.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T28811  
R:Minx, M.

submitted to the EMBL Data Library, January 1996  
A:Description: The sequence of C. elegans cosmid C54D1.  
A:Reference number: Z20527  
A:Accession: T28811  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1557 <MIN>  
A:Cross-references: EMBL:U46673; PIDN:AA048152.1; GSPDB:GNO0028; CESP:C54D1.5  
A:Experimental source: strain Bristol N2; clone C54D1  
C:Genetics:  
A:Gene: CESP:C54D1.5  
A:Map position: X  
A:Introns: 84/3; 127/1; 166/2; 271/2; 331/1; 392/3; 433/2; 585/2; 1089/1; 1530/3  
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology

Query Match 12.4%; Score 68.5; DB 2; Length 1557;  
Best Local Similarity 28.8%; Pred. No. 55;  
Matches 19; Conservative 5; Mismatches 33; Indels 9; Gaps 2;

QY 29 FTVMGPGCNNAERXKCGCAIHOKGCDYDSTGCTALYNQAG--CGVAHTRFSS 86  
DB 920 YNTSGLG-----QECNCDPLSGEVTCDVNTGCCCKEYVGRCDCADYHFGFS 972

QY 87 ARACNP 92  
DB 973 ANGCOP 978

RESULT 6  
DVTU1  
multidrug resistance protein 1 - human  
N:Alternate names: P-glycoprotein 1  
C:Species: Homo sapiens (man)  
C>Date: 31-Dec-1990 #sequence\_revision 18-Aug-1995 #text\_change 19-Jan-2001  
R:Accession: A34914; PS0162; S15500; S43838; I52238; I55204  
R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.  
J. Biol. Chem. 265, 506-514, 1990  
A:Title: Genomic organization of the human multidrug resistance (MDR1) gene and origin  
A:Reference number: A34914; MUID:90094448; PMID:1967175

A:Molecule type: DNA  
A:Residues: 1-1280 <CHE>  
A:Cross-references: GB:M29447; GB:J05168; NID:G187496; PIDN:AAA59576.1; PID:G386862  
R:Kloka, N.; Yamano, Y.; Komano, T.; Ueda, K.  
submitted to JIPID, April 1991  
A:Reference number: PS0162  
A:Accession: PS0162

A:Molecule type: DNA  
A:Residues: 1-22 <KIO>  
R:Kloka, N.; Yamano, Y.; Komano, T.; Ueda, K.  
submitted to the EMBL Data Library, April 1991  
A:Description: Transcriptional regulation of multidrug resistance gene (MDR1) expression  
A:Reference number: S15500  
A:Accession: S15500

A:Molecule type: DNA  
A:Residues: 1-22, 'R' <K12>  
A:Cross-references: EMBL:X58723; NID:G34522; PIDN:CAA1558.1; PID:G34523  
R:Chen, C.; Chin, J.B.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.  
Cell 47, 381-389, 1986  
A:Title: Internal duplication and homology with bacterial transport proteins in the mdr1  
A:Reference number: A25059; MUID:87028230; PMID:2876781  
A:Accession: A25059

A:Molecule type: mRNA  
A:Residues: 1-184, 'V' 186-1280 <CH2>  
A:Cross-references: GB:M14758; NID:G187468; PIDN:AAA59575.1; PID:G307180  
R:Chambers, T.C.; Pohl, J.; Glaeser, D.B.; Kuo, J.F.  
Biochem. J. 299, 309-315, 1994  
A:Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase of  
A:Reference number: S43838; MUID:94220047; PMID:7909431  
A:Accession: S43838

A:Molecule type: protein  
A:Residues: 656-689 <CHA>  
R:Geleler, V.; Weger, S.; Probst, H.

Biochem. Biophys. Res. Commun. 169, 796-802, 1990  
A:Title: mdr1/P-glycoprotein gene segments analyzed from various human leukemic cell lines  
A:Reference number: I52238; MUID:90250529; PMID:1572623  
A:Accession: I52238  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 178-215 <RES>  
A:Cross-references: GB:M37724; NID:G183537; PIDN:AAA88047.1; PID:G553314  
A:Accession: I65204

A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 800-856 <RE2>  
A:Cross-references: GB:M37725; NID:G183538; PIDN:AAA88048.1; PID:G553315  
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell  
structurally and functionally unrelated lipophilic anticancer drugs.  
C:Genetics:  
A:Gene: GDB:PCY1, MDR1  
A:Cross-references: GDB:120712; OMIM:171050  
A:Map position: 7q21-7q21  
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology  
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; phosphoprotein;  
F:1-638,653-1280/Region: duplication  
F:49-350/Domain: hydrophobic <HB1>  
F:351-637/Domain: hydrophilic <HL1>  
F:410-604/Domain: ATP-binding cassette homology <ABC1>  
F:427-434/Region: nucleotide-binding motif A (P-loop)  
F:551-555/Region: nucleotide-binding motif B  
F:638-708/Domain: linker <LIN>  
F:709-993/Domain: hydrophobic <HB2>  
F:994-1280/Domain: hydrophilic <HL2>  
F:1053-1249/Domain: ATP-binding cassette homology <ABC2>  
F:1070-1077/Region: nucleotide-binding motif A (P-loop)  
F:1196-1200/Region: nucleotide-binding motif B  
F:91,94,99/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:433/Binding site: ATP (Lys) #status predicted  
F:661,667/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status ext  
F:667,671,683/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status  
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 12.3%; Score 68; DB 1; Length 1280;  
Best Local Similarity 24.5%; Pred. No. 51;  
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFPSVTVMMLTAMSENV-----NGSAFTWSGPGCNNAERXKCGCAIHOKGCDY 59  
DB 334 VFPSVLIGAFSVGQSPSTEAFANRGAAYEFKILDNKPSIDISKSGHKPDNIKGLDE 393

QY 60 -----FSY-----TGCTALYNQAG 75  
DB 394 FRNVFSPYSRKREVKILKGLNKLKVSQGTVALVNSGC 431

RESULT 7  
T16044  
hypothetical protein F12A10.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
A:Accession: T16044  
R:Geleler, V.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid F12A10.  
A:Reference number: Z18451  
A:Accession: T16044

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-184 <GRT>  
A:Cross-references: EMBL:U28731; NID:G861241; PID:G861242; PIDN:AAA68294.1; CESP:F12A10  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F12A10.1  
A:Introns: 25/2; 66/1, 114/2

Query Match 12.3%; Score 67.5; DB 2; Length 184;

Best Local Similarity 24.1%; Pred. No. 9.6;  
Matches 26; Conservative 11; Mismatches 34; Indels 37; Gaps 4;  
QY 1 MASTKLFFSVITVWMLIAMASEMVN-----GSAFTVWSG-----PG 36  
Db 78 MHSTLFFVVFILVAVSLAFDDLHPKLNQWRSSGSRVKRMGGGYPGGYGGGYG 137  
QY 37 CNNRAERYSKGCSATHOKGYDFSYTGQTAALYNQAGCGVAHTRFG 84  
Db 138 -----GYSGY-----GGYFGYGGGSGYSSSSRSRG 172  
RESULT 8  
T02667  
proteinase inhibitor - rice  
C:Species: Oryza sativa (rice)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jun-2000  
C:Accession: T02667  
R:Yun, C.H.; Lee, J.H.; Park, J.H.; Lee, G.R.  
submitted to the EMBL Data Library, January 1998  
A:Description: Molecular characterization of rice proteinase inhibitor gene.  
A:Reference number: Z14694  
A:Accession: T02667  
A:Status: preliminary; translated from GB/EMBL/DBSJ  
A:Molecule type: DNA  
A:Residues: 1-81 <YUN>  
A:Cross-references: EMBL:AF044059; NID:g2829211; PIDN:AAC00503.1; PID:g2829212  
A:Experimental source: strain IR36  
C:Genetics:  
A:Gene: RGP19  
A:Introns: 21/1  
C:Superfamily: gamma-thionin  
Query Match 12.1%; Score 66.5; DB 2; Length 81;  
Best Local Similarity 23.2%; Pred. No. 5.7;  
Matches 19; Conservative 16; Mismatches 28; Indels 19; Gaps 3;  
QY 2 ASTKLFFSVITVWMLIAMASEMVNAGSAFTVWSGPGCNNAERYS-----KGCSSAIHQK 56  
Db 3 ASRKVSAMLLMWLLAATGEM--GGPWVAEARTCSQSHRFKPGPCARKANCASVCNTE 60  
QY 57 GYDFSYTGQTAALYNQAGCGSV 78  
Db 61 GPPDGY-----CHGV 70  
RESULT 9  
JC6040  
fimbriin protein agfB precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6040  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6040  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43598.1; PID:gl184713  
A:Experimental source: strain 276755-3b  
C:Genetics:  
A:Gene: agfB  
C:Function:  
A:Description: minor component of thin aggregative fimbriae  
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbria  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-151/Product: fimbriin protein agfB #status predicted <MAT>  
Query Match 12.1%; Score 66.5; DB 2; Length 151;  
Best Local Similarity 31.0%; Pred. No. 10;  
Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;

QY 30 TVWSGPGCNNAERYSKGCSATHOKGYDFSYTGQTAALYNQAGCGVAHTRFGSSA 87  
Db 69 SVISOEGGNRAK-----VDQAGNYNFAVTEQT-----GNANDASISQSAYGNSA 113  
RESULT 10  
S31078  
seed allergen RA5 - rice  
C:Species: Oryza sativa (rice)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Jun-2000  
C:Accession: S31078  
R:Adachi, T.; Izumi, H.; Yamada, T.; Tanaka, K.; Takeuchi, S.; Nakamura, R.; Matsuoka, T.  
Plant Mol. Biol. 21, 239-248, 1993  
A:Title: Gene structure and expression of rice seed allergenic proteins belonging to the  
A:Reference number: S31078; MUID:93144699; PMID:7678765  
A:Accession: S31078  
A:Molecule type: mRNA  
A:Residues: 1-157 <ADA>  
A:Cross-references: EMBL:D11430; NID:g218196; PIDN:BAA01996.1; PID:g218197  
C:Superfamily: wheat alpha-amylase inhibitor  
C:Keywords: seed  
Query Match 12.1%; Score 66.5; DB 2; Length 157;  
Best Local Similarity 22.1%; Pred. No. 11;  
Matches 33; Conservative 12; Mismatches 35; Indels 69; Gaps 7;  
QY 1 MASTKLFFSV--ITVWMLIAMASEMVN-----GSAFTVWSGPGCNNAERY 42  
Db 1 MASNKVWFVSVLLLVAVSVLAATATMAEYHQDQVYVTRARCPQGMGYPMYSLPRCALVK 60  
QY 43 RY-----SKGCSAI-HQKGYDFSYTGQTAALYNQAGCGS 77  
Db 61 RQCRGSAAAEQVRDCCRQLAAVDDSMCRCEAISHMLGG-----IYRELGA 109  
QY 78 VAH-----TRFGSSARA-----CN 91  
Db 110 VGHPMSEVPRGCRGDLERAAASLPFCN 138  
RESULT 11  
S21157  
seed allergen RA17 - rice  
C:Species: Oryza sativa (rice)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Jun-2000  
C:Accession: S21157; JC4887; S31079  
R:Izumi, H.; Adachi, T.; Fujii, N.; Matsuoka, T.; Nakamura, R.; Tanaka, K.; Uriasu, A.; Ku:  
FEBS Lett. 302, 213-216, 1992  
A:Title: Nucleotide sequence of a cDNA clone encoding a major allergenic protein in rice  
A:Reference number: S21157; MUID:92289999; PMID:1376283  
A:Accession: S21157  
A:Molecule type: mRNA  
A:Residues: 1-162 <IZU>  
A:Cross-references: EMBL:D11431; NID:g218194; PIDN:BAA01997.1; PID:g218195  
R:Nakamura, R.; Matsuoka, T.  
BioSci. Biotechnol. Biochem. 60, 1215-1221, 1996  
A:Title: Rice allergenic protein and molecular-genetic approach for hypoallergenic rice.  
A:Reference number: JC4887; MUID:97141195; PMID:8987539  
A:Contents: seed  
A:Accession: JC4887  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-162 <NKA>  
A:Cross-references: EMBL:X66257; NID:g311892; PIDN:CAA46983.1; PID:g311893  
C:Comment: This protein shows a sequence similarity to alpha-amylase/trypsin inhibitor f.  
C:Superfamily: wheat alpha-amylase inhibitor  
Query Match 12.1%; Score 66.5; DB 2; Length 162;  
Best Local Similarity 21.5%; Pred. No. 11;  
Matches 31; Conservative 17; Mismatches 37; Indels 59; Gaps 7;  
QY 1 MASTKLFFSVITVWMLIAMASEMVN-----GSAFTVWSGPGCNNAERY 44

6

Db 1 MASNKVSVLLVLSVLAAMATADHHQVYSPGEGCRPGISYPTVSLPOCRITLVR- 59

QY 45 SKC---GCSAIHQK-----GGYDFSYGTOTALYNQAGCSGVH-- 80

Db 60 -QCVGRGASNADEQWQDCRQLAAVDGMCRCGALDHLHLSG-----ITRELGATEAGHPM 114

QY 81 -----TRGSSSRA-----CN 91

Db 115 AEVFPCCRGRGDLERMAASLPAFCN 138

RESULT 12

T34656

hypothetical protein SCIA9.07 SCIA9.07 - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

R/Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.

submitted to the EMBL Data Library, December 1998

A/Accession: T34656

A/Reference number: 221552

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-251 <SAU>

A/Cross-references: EMBL:AL034446; PIDN:CAA2377.1; GSPDB:GN00070; SCQEDB:SCIA9.07

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: SCQEDB:SCIA9.07

Query Match 12.1%; Score 66.5; DB 2; Length 251;

Best Local Similarity 33.3%; Pred. No. 16;

Matches 23; Conservative 4; Mismatches 29; Indels 13; Gaps 3;

QY 24 VNGSAFTWSSGSG---CNRAERYSKGCSAIHQKGYDFSYGTALYN-----LYN 71

Db 134 VNARTDTWSSGSGVETLRLRLRYREAGAGVFPGLTBPARGSLAARFDVPLNLYR 193

QY 72 QAGCSGVH 80

Db 194 PAG-PGLAH 201

RESULT 13

AE0311

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain F [imported] - Yersinia pestis (str

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 03-Jun-2002

R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tildall, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AE0311

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-461 <KUR>

A/Cross-references: GB:AL590842; PIDN:CAK91353.1; PID:G15980542; GSPDB:GN00175

C/Genetics:

A/Gene: nuoc

C/Superfamily: NADH dehydrogenase (ubiquinone) chain F; NADH dehydrogenase (ubiquinone)

C/Keywords: oxidoreductase

Query Match 12.1%; Score 66.5; DB 2; Length 461;

Best Local Similarity 34.8%; Pred. No. 29;

Matches 24; Conservative 3; Mismatches 33; Indels 9; Gaps 3;

QY 31 VNGSGCNRRAERYSKGCSAIHQKGYDFSYGTOTALYNQAGCSGVHTRFGSSARAC 90

Db 216 VMKRPCTVNVNVE--TLCNVPALIEHGV--WVGIRAGKSNAGTK-----LMGFSGRVK 266

QY 91 NPGWMSIF 99

Db 267 NFGMLPLP 275

RESULT 14

DVMS2

multidrug resistance protein 2 - mouse

N/Alternate names: P-glycoprotein MDR2

C/Species: Mus musculus (house mouse)

C/Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 19-Jan-2001

C/Accession: A30409; S70711

R/Gros, P.; Raymond, M.; Bell, J.; Housman, D.

Mol. Cell. Biol. 8, 2770-2778, 1988

A/Title: Cloning and characterization of a second member of the mouse mdr gene family.

A/Reference number: A30409; MUID:88302195; PMID:3405218

A/Accession: A30409

A/Molecule type: mRNA

A/Residues: 1-1276 <HSU>

A/Cross-references: GB:U03398; NID:G199109; PIDN:AAA39516.1; PID:G387428

R/Kirschner, L.S.

Nucleic Acids Res. 24, 2829-2834, 1996

A/Title: De novo generation of simple sequence during gene amplification.

A/Reference number: S70711; MUID:96313253; PMID:8759018

A/Accession: S70711

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 43-92 <KIR>

A/Cross-references: EMBL:U46839; NID:G1228142; PIDN:AAC52722.1; PID:G1228143

A/Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1996

C/Comment: This is an integral membrane protein overproduced in multidrug-resistant cell

culturally and functionally unrelated lipophilic anticancer drugs.

C/Genetics:

A/Gene: mdr2

C/Superfamily: multidrug resistance protein; ATP-binding cassette homology

C/Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane p

F/1-637/653/Domain: Region: duplication

F/409-603/Domain: ATP-binding cassette homology <ABC1>

F/426-433/Region: nucleotide-binding motif A (P-loop)

F/550-554/Region: nucleotide-binding motif B

F/1049-1245/Domain: ATP-binding cassette homology <ABC2>

F/1067-1074/Region: nucleotide-binding motif A (P-loop)

F/1192-1196/Region: nucleotide-binding motif B

F/88/94/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/432/Binding site: ATP (lys) #status predicted

F/1072/Binding site: ATP (lys) #status predicted

Query Match 12.0%; Score 66; DB 1; Length 1276;

Best Local Similarity 20.4%; Pred. No. 82;

Matches 20; Conservative 19; Mismatches 31; Indels 28; Gaps 2;

QY 6 LFSVITVWMLIMASEMVN-----GSAFTWSSGCGCNRRAERYSKGCSAIHQKGYD 59

Db 333 VFSILGASVQGAAPCIDAFANAGAAVYFDIINDNPKIDSFSERGHKPKDNKMLE 392

QY 60 FS-----YTGTALYNQAGC 75

Db 393 FSDVHFSYSPRANIKILKGLNKVSGQTVALVNSGC 430

RESULT 15

AH0635

nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica

C/Species: Salmonella enterica subsp. enterica serovar Typhi

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001

C/Accession: AH0635

R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero

A/Reference number: AB0502; PMID:11677608

us-09-882-434a-1.rpr

Mon Jan 13 09:25:46 2003

A;Accession: AH0635  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-151 <PAR>  
 A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176  
 C;Genetics:  
 A;Gene: STY1180

Query Match 11.9%; Score 65.5; DB 2; Length 151;  
 Best Local Similarity 31.0%; Pred. No. 13;  
 Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;  
 QY 30 TVWSGPCNNRERYSKGCSAIHKGGYDFSYTGQTAAALYNQAGCSGVAHTRFGSSA 87  
 Db 69 SVISOEGENNRK-----VDQAGNYNFAYIEQT-----GNANDASISQSAVGN 113

Search completed: January 12, 2003, 09:39:24  
 Job time : 45 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 08:42:05 ; Search time 25 Seconds

(without alignments)  
169.223 Million cell updates/sec

Title: US-09-882-434a-1

Perfect score: 551  
Sequence: 1 MASTKLFPSVITVMMILMAA.....FGSSARACNPFMGKSIPIQC 102

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	551	100.0	102	1	AMP1 MACIN
2	75.5	13.7	485	1	HXKB YEAST
3	71.5	13.0	403	1	PGK CHLNU
4	69	12.5	1281	1	MDR1 CRIGR
5	68.5	12.4	1557	1	LMU1 CAEBL
6	68	12.3	1280	1	MDR1 HUMAN
7	67.5	12.3	902	1	VEF GVHA
8	66.5	12.1	157	1	CSGB SALTY
9	66.5	12.1	162	1	RA17 ORYSA
10	66.5	12.1	162	1	MDR2 MOUSE
11	66	12.0	1276	1	MDR2 MOUSE
12	65.5	11.9	151	1	CSGB SALTY
13	65.5	11.9	244	1	OSL3 ARATH
14	65	11.8	165	1	RA14 ORYSA
15	64.5	11.7	485	1	HXKA YEAST
16	64	11.6	320	1	RLX1 STAU
17	64	11.6	1278	1	MDR2 RAT
18	64	11.6	1609	1	LMG1 HUMAN
19	63.5	11.5	166	1	RAG2 ORYSA
20	63.5	11.5	513	1	Y4WI RHISN
21	63.5	11.5	828	1	MRKC KLEPN
22	63.5	11.5	1019	1	LFC CARRO
23	63.5	11.5	1019	1	LFC CARRO
24	63	11.4	159	1	PRO1 LYCES
25	63	11.4	1279	1	MDR3 HUMAN
26	62.5	11.3	235	1	THM2 THADA
27	62.5	11.3	250	1	EXR1 ARATH
28	62.5	11.3	488	1	CRYP CRYPE
29	62	11.3	159	1	PRO6 LYCES
30	62	11.3	159	1	HSEC DROME
31	62	11.2	159	1	YB12 YEAST
32	61.5	11.2	455	1	AT11 ARATH
33	61	11.1	501	1	AT11 ARATH

## ALIGNMENTS

RESULT 1	AMP1 MACIN	STANDARD	PRT	102 AA.	
AC	P80915: 004396:				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Antimicrobial peptide 1 precursor (AMP1).				
OS	Macadamia integrifolia (Macadamia nut).				
OC	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.				
OX	NBI_TaxID=60698;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 27-96.				
RC	TISSUE=Seed;				
RA	MEDLINE=97261828; PubMed=9108242;				
RT	Marcus J.P., Green J.L., Goulter K.C., Harrison S.J., Mannens J.M.;				
RT	"Purification, characterisation and cDNA cloning of an antimicrobial				
RL	peptide from Macadamia integrifolia."				
RL	Eur. J. Biochem. 244:743-749(1997).				
CC	-1- FUNCTION: ANTIMICROBIAL PEPTIDE WHICH INHIBITS THE GROWTH OF				
CC	A VARIETY OF FUNGI, OOMYCETES, GRAM-POSITIVE BACTERIAL				
CC	PHYTOPATHOGENES AND S.CEREVISIAE IN VITRO. NO ACTIVITY AGAINST				
CC	E.COLI.				
CC	-1- MISCELLANEOUS: ITS ANTIMICROBIAL ACTIVITY IS DIMINISHED BY CALCIUM				
CC	AND POTASSIUM CHLORIDE SALTS.				
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib.ch).				
DR	EMBL, Y10903; CAA71842.1; -				
KW	Plant defense; Antibiotic; Fungicide; Signal.				
FT	SIGNAL				
FT	CHAIN 1 27 102				
SQ	SEQUENCE 102 AA; 10943 MW; 9FB799544549311 CRC64;				

Query Match 100.0%; Score 551; DB 1; Length 102;  
Best Local Similarity 100.0%; Pred. No. 1.0e-51;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASTKLFPSVITVMMILMAAEMVNGSAFTVWSGCGCNRAERYSKGCSAIIHOKGVDF	60
DB	1	MASTKLFPSVITVMMILMAAEMVNGSAFTVWSGCGCNRAERYSKGCSAIIHOKGVDF	60
QY	61	STTGQTALYNQAGCGVAHTRFGSSARACNPFMGKSIPIQC	102
DB	61	STTGQTALYNQAGCGVAHTRFGSSARACNPFMGKSIPIQC	102

RESULT 2  
HXKB YEAST



FT TURN 165 166  
 FT STRAND 169 169  
 FT TURN 175 176  
 FT STRAND 187 187  
 FT HELIX 188 199  
 FT STRAND 202 208  
 FT HELIX 210 221  
 FT TURN 223 224  
 FT STRAND 225 231  
 FT STRAND 235 241  
 FT HELIX 243 245  
 FT TURN 251 252  
 FT STRAND 263 266  
 FT TURN 270 275  
 FT HELIX 283 291  
 FT HELIX 299 304  
 FT HELIX 306 322  
 FT TURN 323 324  
 FT TURN 338 339  
 FT TURN 343 343  
 FT HELIX 344 351  
 FT HELIX 358 368  
 FT TURN 369 369  
 FT HELIX 374 395  
 FT TURN 395 396  
 FT HELIX 397 406  
 FT TURN 407 407  
 FT STRAND 411 416  
 FT TURN 418 421  
 FT TURN 424 425  
 FT HELIX 426 438  
 FT HELIX 445 447  
 FT STRAND 450 454  
 FT TURN 458 460  
 FT HELIX 461 475  
 SQ SEQUENCE 485 AA; 53811 MW; E4A7EC796125324E CRC64;

Query Match 13.7%; Score 75.5; DB 1; Length 485;  
 Best Local Similarity 30.2%; Pred. No. 1;  
 Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;

QY 8 FSATYVMM---LIMASEMNGSAFTVWSPGCGNNRERYSKCSAIIHOKGSDPSYGT 64  
 DB 368 FGINTTYOEKRLIRRELGA-----RAARLSVGIATICQKRGK---TG 411  
 QY 65 QTAALYNQAGCSGVATRF-GSSARACNP---FGW 95  
 DB 412 HIAA-----DGSVYRYPGPEKAKANALKDLYGW 440

RESULT 3  
 ID PGK CHLMU STANDARD; PRT; 403 AA.  
 AC 09PLN4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phosphoglycerate kinase (EC 2.7.2.3).  
 GN PGK OR TC0065.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NC NCBI\_TaxID=83560;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / N199;  
 RA MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,  
 Linher K., Weidman J., Knout H., Craven B., Bowman C., Dodson R.,  
 RA Guin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
 pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-  
 CC phospho-D-glyceroyl phosphate.  
 CC -1- PATHWAY: Second phase of glycolysis; second step.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE002274; AAF73528.1; -  
 DR HSRP: P36204; IYPE.  
 DR TIGR: TC0065; -  
 DR InterPro: IPR001576; PGK.  
 DR Pfam: PF00162; PGK; 1.  
 DR PRINTS: PRO0477; PHGLYCKINASE.  
 DR PROSITE: PS00111; PGLYCERATE KINASE; 1.  
 KW Transferase; kinase; glycolysis; Complete proteome.  
 SQ SEQUENCE 403 AA; 43391 MW; 7A80C7A55089F64 CRC64;

Query Match 13.0%; Score 71.5; DB 1; Length 403;  
 Best Local Similarity 29.3%; Pred. No. 2.3;  
 Matches 24; Conservative 13; Mismatches 28; Indels 17; Gaps 4;

QY 17 IMASEMNGSAFTVWSC-----PGCNRERYSKCSAIIHOKGSDPSYGT 66  
 DB 307 IALFSEVIODSATVFMNPGVGEVPPDQSKAIAQCLASHSAVTVWG-----GDA 360  
 QY 67 AALYNQAGC-SGVATRFSSA 87  
 DB 361 AAVALAGCTSQISHVSTGGG 382

RESULT 4  
 ID MDR3 CRIGR STANDARD; PRT; 1281 AA.  
 AC P23174;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Multidrug resistance protein 3 (P-glycoprotein 3).  
 GN PGY3 OR PGP3.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.  
 NC NCBI\_TaxID=10029;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=92135896; PubMed=1685679;  
 RA Endicott J.A., Sarangi F., Ling V.;  
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein  
 RT gene family";  
 RL DNA Seq. 2:89-101(1991).  
 CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED  
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT  
 CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY  
 CC CANNOT.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M60042; AAA68885.1; .  
 DR HSP; P13569; INBD.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR001140; ABCtransporter.  
 DR Pfam; PF00005; ABC tran; 2.  
 DR Pfam; PF00664; ABC membrane; 2.  
 DR ProDom; PD000006; ABC transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC TRANSPORTER; 1.  
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;  
 KW Multigene family.  
 FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 58 78 POTENTIAL.  
 FT TRANSMEM 122 142 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 218 238 POTENTIAL.  
 FT TRANSMEM 299 319 POTENTIAL.  
 FT TRANSMEM 328 348 POTENTIAL.  
 FT DOMAIN 349 712 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 713 733 POTENTIAL.  
 FT TRANSMEM 758 778 POTENTIAL.  
 FT TRANSMEM 834 854 POTENTIAL.  
 FT TRANSMEM 855 875 POTENTIAL.  
 FT TRANSMEM 938 958 POTENTIAL.  
 FT TRANSMEM 975 995 POTENTIAL.  
 FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 429 436 ATP (POTENTIAL).  
 FT NP\_BIND 1071 1078 ATP (POTENTIAL).  
 SQ SEQUENCE 1281 AA; 140866 MW; 2203EF61EBB29602 CRC64;  
 Query Match 12.5%; Score 69; DB 1; Length 1281;  
 Best Local Similarity 21.4%; Pred. No. 12;  
 Matches 2; Conservative 17; Mismatches 32; Indels 28; Gaps 2;  
 QY 6 LFFSVITVMMIIMASEMVN-----GSAFTVSGPGCNRNRAERYKCGCSATHQKGGYD 59  
 DB 336 VFFSILIGAFSGVQAAPCIDAFANARGAAVYFIIDNNPKIDFSRGRHKPSIKGNLD 395  
 QY 60 FS-----YTGOTALYNQAGC 75  
 DB 396 FSDVHFSPSRANIKILKGLNLKLVQSGGTVALVNSGC 433  
 RESULT 5  
 LML1\_CAEEL STANDARD; PRT; 1557 AA.  
 AC Q18823;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Laminin-like protein C54D1.5 precursor.  
 GN C54D1.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Minx P.;  
 RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

EMBL; U46673; AAC48152.1; .  
 HSP; P02468; 1TLE.  
 WormPep; C54D1.5; CE06981.  
 InterPro; IPR000561; EGF-like.  
 InterPro; IPR001886; LamNT.  
 InterPro; IPR000034; Laminin\_B.  
 InterPro; IPR002049; Laminin\_EGF.  
 Pfam; PF00052; laminin\_B; 1.  
 Pfam; PF00053; laminin\_EGF; 10.  
 Pfam; PF00055; laminin\_Nterm; 1.  
 PRINTS; PR00011; EGFLAMININ.  
 ProDom; PD002082; LamNT; 1.  
 ProDom; PD003031; Laminin\_B; 1.  
 SMART; SM00180; EGF Lam; 9.  
 SMART; SM00001; EGF-like; 2.  
 SMART; SM00281; LamB; 1.  
 SMART; SM00136; LamNT; 1.  
 PROSITE; PS00022; EGF\_1; 8.  
 PROSITE; PS01186; EGF\_2; 1.  
 PROSITE; PS01248; LAMININ TYPE EGF; 11.  
 KW Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.  
 FT SIGNAL 1 1557 POTENTIAL.  
 FT CHAIN ? 1557 LAMININ-LIKE PROTEIN C54D1.5.  
 FT DOMAIN ? 271 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 272 331 LAMININ EGF-LIKE 1.  
 FT DOMAIN 332 387 LAMININ EGF-LIKE 2.  
 FT DOMAIN 388 434 LAMININ EGF-LIKE 3.  
 FT DOMAIN 435 487 LAMININ EGF-LIKE 4.  
 FT DOMAIN 488 497 LAMININ EGF-LIKE 5 (N-TERMINAL).  
 FT DOMAIN 498 688 LAMININ DOMAIN IV.  
 FT DOMAIN 687 732 LAMININ EGF-LIKE 5 (C-TERMINAL).  
 FT DOMAIN 737 769 LAMININ EGF-LIKE 6 (INCOMPLETE).  
 FT DOMAIN 770 819 LAMININ EGF-LIKE 7.  
 FT DOMAIN 820 874 LAMININ EGF-LIKE 8.  
 FT DOMAIN 875 930 LAMININ EGF-LIKE 9.  
 FT DOMAIN 931 978 LAMININ EGF-LIKE 10.  
 FT DOMAIN 979 1025 LAMININ EGF-LIKE 11.  
 FT DISULFID 272 281 BY SIMILARITY.  
 FT DISULFID 274 295 BY SIMILARITY.  
 FT DISULFID 297 306 BY SIMILARITY.  
 FT DISULFID 309 329 BY SIMILARITY.  
 FT DISULFID 332 341 BY SIMILARITY.  
 FT DISULFID 334 357 BY SIMILARITY.  
 FT DISULFID 360 369 BY SIMILARITY.  
 FT DISULFID 372 385 BY SIMILARITY.  
 FT DISULFID 388 400 BY SIMILARITY.  
 FT DISULFID 390 406 BY SIMILARITY.  
 FT DISULFID 408 417 BY SIMILARITY.  
 FT DISULFID 420 432 BY SIMILARITY.  
 FT DISULFID 435 449 BY SIMILARITY.  
 FT DISULFID 437 456 BY SIMILARITY.  
 FT DISULFID 458 467 BY SIMILARITY.  
 FT DISULFID 470 485 BY SIMILARITY.  
 FT DISULFID 770 778 BY SIMILARITY.  
 FT DISULFID 772 788 BY SIMILARITY.  
 FT DISULFID 791 800 BY SIMILARITY.  
 FT DISULFID 803 817 BY SIMILARITY.  
 FT DISULFID 820 834 BY SIMILARITY.  
 FT DISULFID 822 841 BY SIMILARITY.  
 FT DISULFID 844 853 BY SIMILARITY.  
 FT DISULFID 856 872 BY SIMILARITY.  
 FT DISULFID 875 894 BY SIMILARITY.  
 FT DISULFID 877 901 BY SIMILARITY.  
 FT DISULFID 903 912 BY SIMILARITY.  
 FT DISULFID 915 928 BY SIMILARITY.  
 FT DISULFID 931 943 BY SIMILARITY.  
 FT DISULFID 933 950 BY SIMILARITY.  
 FT DISULFID 952 961 BY SIMILARITY.  
 FT DISULFID 964 976 BY SIMILARITY.

FT DISULFID 979 991 BY SIMILARITY.  
FT DISULFID 981 998 BY SIMILARITY.  
FT DISULFID 1000 1009 BY SIMILARITY.  
FT DISULFID 1012 1023 BY SIMILARITY.  
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1168 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1211 1211 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1557 AA; 172723 MW; CAF0B51F8D558DZF CRC64;  
  
Query Match 12.4%; Score 68.5; DB 1; Length 1557;  
Best Local Similarity 28.8%; Pred. No. 16;  
Matches 19; Conservative 5; Mismatches 33; Indels 9; Gaps 2;  
  
QY 29 FTWSPGCGNNRERYSKCGSAIHQKGYDFSYTGOTALYNQAG--CSGVNHTFGSS 86  
Db 920 YNTSGLG-----QECNCDPLSGEGNTCVNTGCGCKPVGTCRCDCADYHFGFS 972  
QY 87 ARACNP 92  
Db 973 ANCGP 978  
  
RESULT 6  
MDR1 HUMAN  
ID MDR1 HUMAN STANDARD; PRT; 1280 AA.  
AC P08183; 012755; 014812;  
DT 01-ANG-1988 (Rel. 08, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Multidrug resistance protein 1 (P-glycoprotein 1) (CD243 antigen).  
GN ABCB1 OR PGY1 OR MDR1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN 1  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87028230; PubMed=2876781;  
RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,  
RA "Internal duplication and homology with bacterial transport proteins  
RT in the mdr1 (P-glycoprotein) gene from multidrug-resistant human  
RT cells";  
RL Cell 47:381-389(1986).  
RN 12  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90094448; PubMed=1967175;  
RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,  
RA "Genomic organization of the human multidrug resistance (MDR1) gene  
RT and origin of P-glycoproteins";  
RL J. Biol. Chem. 265:506-514(1990).  
RN 13  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97190336; PubMed=9038218;  
RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffeuz J.P.,  
RA Dumontet C., Skic B.I.;  
RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,  
RT altered phenotype, and resistance to cyclosporins";  
RL J. Biol. Chem. 272:5974-5982(1997).  
RN 14  
RP SEQUENCE OF 1-234 FROM N.A.  
RA Smith A., Beck C., Gibson A.;

RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.  
RX MEDLINE=90290529; PubMed=1972623;  
RA Gekeler V., Weger S., Probst H.;  
RT "mdr1/P-glycoprotein gene segments analyzed from various human  
RT leukemic cell lines exhibiting different multidrug resistance  
RT profiles";  
RL Biochem. Biophys. Res. Commun. 169:796-802(1990).  
RN [6]  
RP SEQUENCE OF 1-23 FROM N.A.  
RA Kikwa N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,  
RA Pastan I., Ueda K.;  
RT Submitted (Jul-1991) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP VARIANTS SER-893 AND THR-893.  
RX MEDLINE=21686803; PubMed=11829140;  
RA Saito S., Iida A., Sekine A., Mura Y., Ogawa C., Kawachi S.,  
RA Higuchi S., Nakamura Y.;  
RT "Three hundred twenty-six genetic variations in genes encoding nine  
RT members of ATP-binding cassette, subfamily B (ABCB/MDR/TAP), in the  
RT Japanese population";  
RL J. Hum. Genet. 47:38-50(2002).  
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED  
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE  
CC AND BRAIN.  
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
CC -!- DATABASE: NAME=Atlas Gene. Cytogenet. Oncol. Haematol.;  
CC WWW="http://www.infobiogen.fr/services/chronocancer/genes/pgy1id105.html".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M14758; AAA59575.1; -;  
DR EMBL; M29447; AAA59576.1; -;  
DR EMBL; M29424; AAA59576.1; JOINED.  
DR EMBL; M29425; AAA59576.1; JOINED.  
DR EMBL; M29426; AAA59576.1; JOINED.  
DR EMBL; M29427; AAA59576.1; JOINED.  
DR EMBL; M29428; AAA59576.1; JOINED.  
DR EMBL; M29429; AAA59576.1; JOINED.  
DR EMBL; M29430; AAA59576.1; JOINED.  
DR EMBL; M29432; AAA59576.1; JOINED.  
DR EMBL; M29433; AAA59576.1; JOINED.  
DR EMBL; M29434; AAA59576.1; JOINED.  
DR EMBL; M29435; AAA59576.1; JOINED.  
DR EMBL; M29436; AAA59576.1; JOINED.  
DR EMBL; M29437; AAA59576.1; JOINED.  
DR EMBL; M29438; AAA59576.1; JOINED.  
DR EMBL; M29439; AAA59576.1; JOINED.  
DR EMBL; M29440; AAA59576.1; JOINED.  
DR EMBL; M29441; AAA59576.1; JOINED.  
DR EMBL; M29442; AAA59576.1; JOINED.  
DR EMBL; M29443; AAA59576.1; JOINED.  
DR EMBL; M29444; AAA59576.1; JOINED.  
DR EMBL; M29445; AAA59576.1; JOINED.  
DR EMBL; M29446; AAA59576.1; JOINED.  
DR EMBL; AF016535; AAB69423.1; -;  
DR EMBL; AC002457; AAC82531.1; -;  
DR EMBL; M37724; AAA88047.1; -;  
DR EMBL; M37725; AAA88048.1; -;  
DR EMBL; X59723; CAA41558.1; -;  
DR PIR; A25059; DVH01.  
DR PIR; A34914; A34914.  
DR Genew; HGNC:40; ABCB1.



RA Courtney L., Portojlik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT7.";  
 RL Nature 413:852-856(2001).  
 RP [3]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES=*S. enteritidis*; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
 RT "Salmonella enteritidis *agfBAC* operon encoding thin, aggregative  
 RT fimbriae.";  
 RL J. Bacteriol. 178:662-667(1996).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC -----  
 CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ002301; CAA05316.1; -  
 DR EMBL: AE008749; AAL20073.1; -  
 DR EMBL: U43280; AAC43598.1; -  
 DR StryGene; SG10609; csgb.  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;  
 Query Match 12.1%; Score 66.5; DB 1; Length 151;  
 Best Local Similarity 31.0%; Pred. No. 3.3;  
 Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;  
 Oy 30 TWMSGPCNNRAREYKSCGSAIKQKGYDFSYTGQTALYNQAGCGVAHTFGSSA 87  
 Db 69 SVISQEGNNRAK-----VDQAGNYPAYIEQT-----GNNDASISQAYGNSA 113  
 RESULT 9  
 RA05\_ORYSA STANDARD; PRT; 157 AA.  
 ID RA05\_ORYSA  
 AC Q01861;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Seed allergenic protein RA5 precursor.  
 GN RA5.  
 OS *Oryza sativa* (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriocaricaceae; Oryzaceae; Oryza.  
 NCBI\_Taxid=4530;  
 RX MEDLINE=93144699; PubMed=7678765;  
 RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,  
 RA Nakamura R., Matsuda T.;  
 RT "Gene structure and expression of rice seed allergenic proteins  
 RT belonging to the alpha-amylase/trypsin inhibitor family.";  
 RL Plant Mol. Biol. 21:239-248(1993).  
 CC -1- PTM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR

CC FAMILY.  
 CC -----  
 CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D11430; BAA01996.1; -  
 DR PIR: S31078; S31078.  
 DR HSSP; P01085; IHSS.  
 DR InterPro; IPR003612; AAI.  
 DR InterPro; IPR001768; Try/amy1 inhbr.  
 DR Pfam; PF00234; tryp alpha amyl. 1.  
 DR PRINTS; PR00808; AMLASEINHBTR.  
 DR SMART; SM00499; AAI, 1.  
 DR PROSITE; PS00426; CEREAL\_TRYP\_AMYL\_INH, 1.  
 KW Allergen; Multigene family; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 157 SEED ALLERGENIC PROTEIN RA5.  
 SQ SEQUENCE 157 AA; 17118 MW; C0A5495FFB399E6 CRC64;  
 Query Match 12.1%; Score 66.5; DB 1; Length 157;  
 Best Local Similarity 22.1%; Pred. No. 3.4;  
 Matches 33; Conservative 12; Mismatches 35; Indels 69; Gaps 7;  
 Oy 1 MASTKLFPSV--ITVMMLIMASEMVN-----GSAFTWMSGPCNNRAE 42  
 Db 1 MASNKVFESVLLAVAVSLATATMAEYHODQVYTRACQPGMGWYSLPRRALVK 60  
 Oy 43 RY-----SKGCSAI-HQKGYDFSYTGQTALYNQAGCSG 77  
 Db 61 RQCRGSAABAEQVRDCCRGQLAAVDSSWCRCEATSHMLGCG-----IYRELGAPD 109  
 Oy 78 VAH-----TRFGSAPA-----CN 91  
 Db 110 VGHPMSEVFRGCRGDLBRRAASLPAPCN 138  
 RESULT 10  
 RA17\_ORYSA STANDARD; PRT; 162 AA.  
 ID RA17\_ORYSA  
 AC Q01883;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 28, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Seed allergenic protein RA17 precursor.  
 GN RA17.  
 OS *Oryza sativa* (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriocaricaceae; Oryzaceae; Oryza.  
 NCBI\_Taxid=4530;  
 RX MEDLINE=93144699; PubMed=7678765;  
 RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,  
 RA Nakamura R., Matsuda T.;  
 RT "Gene structure and expression of rice seed allergenic proteins  
 RT belonging to the alpha-amylase/trypsin inhibitor family.";  
 RL Plant Mol. Biol. 21:239-248(1993).  
 CC [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Seed;  
 RC TISSUE=Seed;  
 RX MEDLINE=92289999; PubMed=1376283;  
 RA Izumi H., Adachi T., Fujii N., Matsuda T., Nakamura R., Tanaka K.,  
 RA Urisu A., Kurokawa Y.;  
 RT "Nucleotide sequence of a cDNA clone encoding a major allergenic  
 RT protein in rice seeds. Homology of the deduced amino acid sequence  
 RT with members of alpha-amylase/trypsin inhibitor family.";

RL FBBS Lett. 302:213-216(1992).  
CC -!- PTM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X66257; CAA46983.1; -;  
DR EMBL; D11431; BAA01997.1; -;  
DR PIR; S21157; S21157.  
DR HSSP; P01085; IHSS.  
DR InterPro; IPR003612; AAI.  
DR InterPro; IPR001768; Try/alpha inhbr.  
DR Pfam; PF00234; try\_alpha\_ami1; 1.  
DR PRINTS; PR00808; AMLASEINHBT.  
DR SMART; SM00499; AAI; 1.  
DR PROSITE; PS00426; CEREAL TRYP AMYL INH; 1.  
KW Allergen; Multigene family; Signal-  
FT SIGNAL 1 27 POTENTIAL  
FT CHAIN 28 162 SEED ALLERGENIC PROTEIN RA17.  
FT SEQUENCE 162 AA; 17497 MW; EA5CDE021FBA9348 CRC64;  
SQ  
Query Match 12.1%; Score 66.5; DB 1; Length 162;  
Best Local Similarity 21.5%; Pred. No. 3.5;  
Matches 31; Conservative 17; Mismatches 37; Indels 59; Gaps 7;  
Qy 1 MASTKLFVSIVTMMLIAMASEMVN-----GSAFTVSGPGCNRAERY 44  
Db 1 MASNKVFSVLLLVSLVLAAMATMADHHQVSGPGCRPGISYPTSLPQCRRLVRR- 59  
Qy 45 SKC---CCSAIHQK-----GGYDFSVTGTQTAALYNQAGCGVAH-- 80  
Db 60 -QCVGRGASAADEQVQDCRCQLAADVGGWCRGALDHMLSG-----IYRELGATEAGHPM 114  
Qy 81 -----TRFGSSARA-----CN 91  
Db 115 AEVPGCGRGDLERAASLPAFCN 138  
RESULT 11  
MDR2 MOUSE  
ID MDR2\_MOUSE STANDARD; PRT; 1276 AA.  
AC P21440;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Multidrug resistance protein 2 (P-glycoprotein 2).  
GN ABCB4 OR PGY2 OR PGV-2 OR MDR2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88302195; PubMed=3405218;  
RA Gros P., Raymond M., Bell J., Housman D.;  
RT "Cloning and characterization of a second member of the mouse mdr  
gene family";  
RL Mol. Cell. Biol. 8:2770-2778(1988).  
RN [2]  
RP SEQUENCE OF 1-23 FROM N.A.  
RC STRAIN=BALE/c;  
RA Kirschner L.S., Horwitz S.B.;  
RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED  
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. MOUSE MDR2 IS  
CC NOT CAPABLE OF CONFERRING DRUG RESISTANCE.

CC  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE  
CC RELATED BUT DISTINCT CELLULAR GENES.  
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; J03398; AAA39516.1; -;  
DR EMBL; M74151; AAA39515.1; -;  
DR PIR; A30409; DVMS2.  
DR HSSP; P13569; INBD.  
DR MGD; MGI:97569; Abcb4.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR003439; ABC\_transportr.  
DR InterPro; IPR001140; ABCtransportr.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR Pfam; PF00664; ABC\_membrane; 2.  
DR ProDom; PD000006; ABC\_transportr; 2.  
DR SMART; SM00382; AAA; 2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 2.  
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;  
KW Multigene family.  
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 52 75 POTENTIAL.  
FT TRANSMEM 119 139 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 215 235 POTENTIAL.  
FT TRANSMEM 296 316 POTENTIAL.  
FT TRANSMEM 325 345 POTENTIAL.  
FT DOMAIN 346 707 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 708 728 POTENTIAL.  
FT TRANSMEM 753 773 POTENTIAL.  
FT TRANSMEM 829 849 POTENTIAL.  
FT TRANSMEM 850 870 POTENTIAL.  
FT TRANSMEM 933 953 POTENTIAL.  
FT TRANSMEM 970 990 POTENTIAL.  
FT DOMAIN 991 1276 CYTOPLASMIC (POTENTIAL).  
FT NP\_BIND 426 433 ATP (BY SIMILARITY).  
FT NP\_BIND 1066 1073 ATP (BY SIMILARITY).  
FT REPEAT 1 635  
FT REPEAT 636 1276  
FT SEQUENCE 1276 AA; 140332 MW; A6C38DCD2C118EEF CRC64;  
SQ  
Query Match 12.0%; Score 66; DB 1; Length 1276;  
Best Local Similarity 20.4%; Pred. No. 25;  
Matches 20; Conservative 19; Mismatches 31; Indels 28; Gaps 2;  
Qy 6 LFFSVITVMMLIAMASEMVN-----GSAFTVSGPGCNRAERYSKCGSAIHQGYD 59  
Db 333 VFSSILIGAFVSGQAAPCIDAFANARGAAYVIFDIIDNNPKIDSPSRGHKPKDNKGNLE 392  
Qy 60 FS-----YGTQTAALYNQAGC 75  
Db 393 FSDVHFVSPSRANIKILKGLNLKVKSGQTVALVGNSGC 430  
RESULT 12  
CSGB\_SALTI  
ID CSGB\_SALTI STANDARD; PRT; 151 AA.  
AC Q827M3;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Minor curlin subunit precursor.  
GN CSGB OR Styl180.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;





```
FT DISULFID 143 195 BY SIMILARITY.
FT DISULFID 151 161 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 175 182 BY SIMILARITY.
FT CONFLICT 186 186 E -> V (IN REF. 1).
SQ SEQUENCE 244 AA; 26633 MW; 9FBE9A45E9E195E0 CRC64;

Query Match 11.9%; Score 65.5; DB 1; Length 244;
Best Local Similarity 23.1%; Pred. No. 6.5;
Matches 24; Conservative 19; Mismatches 44; Indels 17; Gaps 3;

QY 1 MASTKLFPSVITVWMLIAMASEMWN--SGPCNNRAERYSKGCSAIHQKGGY 58
DQ 5 LVSTPFSALLIITATATTEILNQCSYTVAAASPGGRRLD-----AGQSRL 55
DQ 59 DFSYTGTAALYNOAGCSGVAHTRFGSSARACNPFQWKSIFQIC 102
DQ 56 DVAAGTKVARIWGRNCRN-----FDSSGRGRCQTDGDCSGGLQC 93

RESULT 14
RA14 ORYZA STANDARD; PRT; 165 AA.
AC Q01882;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Seed allergenic protein RA14 precursor.
GN RA14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=93144699; PubMed=7678765;
RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,
RA Nakamura R., Matsuda T.;
RT "Gene structure and expression of rice seed allergenic proteins
RT belonging to the alpha-amylase/trypsin inhibitor family.";
PL Plant Mol. Biol. 21:239-248(1993).
CC -1- PWM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR ENBL; D11432; BAA01998.1; -.
DR PIR; S31080; S31080.
DR HSP; P01085; IHSS.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Try/amy1 inhbr.
DR Pfam; PF00234; tryp_alpha_aml_1.
DR PRINTS; PR00808; AMLASEINHBR.
DR SMART; SMO0499; AAI; 1.
DR PROSITE; PS00426; CEREAL_TRYP_AMYL_INH; 1.
KW Allergen; Multigene family; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 165 SEED ALLERGENIC PROTEIN RA14.
SQ SEQUENCE 165 AA; 17887 MW; 13F99783726CEFC6A CRC64;

Query Match 11.8%; Score 65; DB 1; Length 165;
Best Local Similarity 20.4%; Pred. No. 5.1;
Matches 31; Conservative 17; Mismatches 32; Indels 72; Gaps 7;
```

```
QY 1 MASTKLFPS--VITVWMLIAMASEMWN-----GSAFTWSPGCGNNRAE 42
DQ 1 MASNKVVSALLIIVSLAATRMADHKKDQVYSLGERCPQNGYPMYSLPRCAVVK 60
QY 43 RY-----SKGCSAI-HOKGGYDFSYTGQTAALYNOAG 74
DQ 61 RQCVGTRSPGAVDEQLAQCCRELAAVDDSWCRCSALNHMVG-----IYRELG 109
QY 75 CSGVAH-----TRFGSSARA-----CN 91
DQ 110 ATDVGHFPAEVPFGCRGDLERAAASLPAFCN 141

RESULT 15
HXKA YEAST STANDARD; PRT; 485 AA.
AC P04806;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hexokinase A (EC 2.7.1.1) (Hexokinase PI).
GN HXK1 OR HKA OR YPR053C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86120382; PubMed=3003701;
RA Stachel C., Stachel J., Swan J., Botstein D., Konigsberg W.;
RT "Identification, cloning and sequence determination of the genes
RT specifying hexokinase A and B from yeast.";
RL Nucleic Acids Res. 14:945-963(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86083199; PubMed=3908224;
RA Kopeck E., Etian K.-D., Mecke D.;
RT "Complete nucleotide sequence of the hexokinase PI gene (HXK1) of
RT Saccharomyces cerevisiae.";
RL Gene 39:95-102(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=96287652; PubMed=9686379;
RA Eki T., Naitou M., Hagiwara H., Ozawa M., Sasanuma S.-I.,
RA Sasanuma M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
RT "Analysis of a 36.2 kb DNA sequence including the right telomere of
RT chromosome VI from Saccharomyces cerevisiae.";
RL Yeast 12:149-167(1996).
RN [5]
RP ATP-BINDING AND SEQUENCE OF 104-112.
RX MEDLINE=88227998; PubMed=3131329;
RA Tamura J.K., Ladime J.R., Cross R.L.;
RT "The adenine nucleotide binding site on yeast hexokinase PII.
RT Affinity labeling of Lys-111 by pyridoxal
RT 5'-diphospho-5'-adenosine.";
RL J. Biol. Chem. 263:7907-7912(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
RX MEDLINE=81049624; PubMed=7001031;
RA Bennett W.S. Jr., Steitz T.A.;
RT "Structure of a complex between yeast hexokinase A and glucose. I.
RT Structure determination and refinement at 3.5-A resolution.";
```

```

RL J. Mol. Biol. 140:183-210(1980).
CC -1- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.
CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
CC -1- INHIBITION BY ATP.
CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBUNIT: HOMODIMER.
CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
CC -1- ISOMERISOMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC -1- DATABASE: NAME=worthington-biochem.com/manual/H/HK.html".
CC WWW="http://www.worthington-biochem.com/manual/H/HK.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14410; AAA34698.1; -.
DR EMBL; X03482; CAA27202.1; -.
DR EMBL; D50617; BAA09292.1; -.
DR PIR; A24531; KIBYHA.
DR PIR; A28178; A28178.
DR PDB; 1HKG; 15-OCT-91.
DR SWISS-2DPAGE; P04806; YEAST.
DR SGD; S0001949; HKX1.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR Pfam; PF03727; hexokinase2; 1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
KW Transferase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;
KW 3D-structure.
KM
FT BINDING 111 111
FT DOMAIN 152 178
FT CONFLICT 61 178
FT CONFLICT 103 103
FT CONFLICT 194 194
FT CONFLICT 244 244
FT CONFLICT 356 357
FT CONFLICT 364 364
FT CONFLICT 388 388
FT CONFLICT 444 444
FT CONFLICT 479 480
FT TURN 19 20
FT HELIX 21 29
FT HELIX 32 33
FT HELIX 38 44
FT TURN 45 46
FT HELIX 47 54
FT TURN 55 56
FT STRAND 66 67
FT STRAND 82 82
FT STRAND 85 85
FT STRAND 88 88
FT TURN 89 90
FT STRAND 91 92
FT STRAND 94 94
FT STRAND 97 97
FT STRAND 111 111
FT TURN 115 119
FT STRAND 121 122
FT TURN 124 126
FT TURN 128 130
FT HELIX 131 140
FT HELIX 142 143
FT STRAND 161 161
FT STRAND 171 171
FT TURN 176 177
FT STRAND 180 181

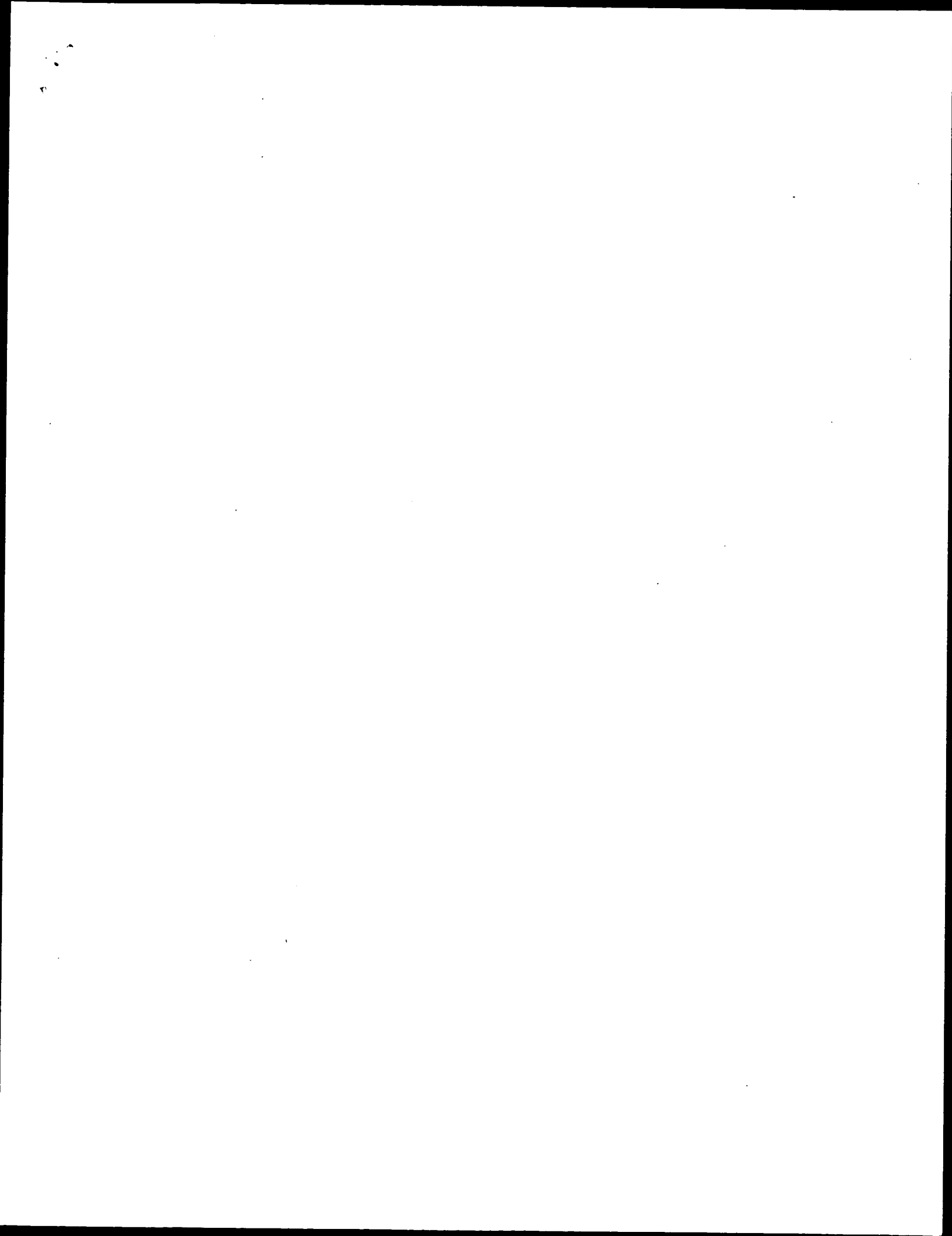
```

```

FT TURN 189 195
FT HELIX 196 200
FT TURN 211 216
FT HELIX 218 222
FT TURN 224 225
FT STRAND 226 232
FT STRAND 236 239
FT STRAND 242 242
FT TURN 244 245
FT STRAND 254 264
FT STRAND 266 267
FT HELIX 270 272
FT TURN 286 287
FT HELIX 288 292
FT TURN 293 294
FT HELIX 300 303
FT TURN 304 305
FT TURN 308 310
FT HELIX 311 323
FT TURN 324 325
FT TURN 339 340
FT TURN 345 351
FT HELIX 359 369
FT TURN 375 377
FT HELIX 378 396
FT TURN 398 401
FT HELIX 402 406
FT TURN 407 408
FT STRAND 412 417
FT TURN 419 420
FT HELIX 428 438
FT STRAND 450 450
FT STRAND 454 454
FT TURN 458 460
FT HELIX 461 474
FT TURN 475 477
SQ SEQUENCE 485 AA; 53738 MW; AF5C9DA8F17BC3D0 CRC64;
Query Match 11.7%; Score 64.5; DB 1; Length 485;
Best Local Similarity 27.4%; Pred. No. 15;
Matches 17; Conservative 9; Mismatches 19; Indels 17; Gaps 2;
QY 40 RAERYKCGSAIHQKGYDPSYTGOTAAIYN-----QGCAGVAHTRFGSSARACNPF 93
DB 391 RAARLAVCGIALICQKRGYKGTGHADGSVYKYPGFKKAARGLR-----DIY 439
QY 94 GW 95
DB 440 GW 441

```

Search completed: January 12, 2003, 09:36:39  
 Job time : 27 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:30:05 ; Search time 80 Seconds

(without alignments)  
262.710 Million cell updates/sec

Title: US-09-882-434A-1

Sequence: 1 MASTKLEFSPVITVMMLIAMA.....FGSSARACNPFMKSLFIQC 102

Scoring table:

BIOSUN62

Gap: 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP yvirus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.5	14.1	486	3	Q05838 saccharomyc
2	75.5	13.7	1558	5	Q9BLR5 Q9BLR5
3	75	13.6	254	5	Q45025 hydra magni
4	71	12.9	730	2	Q07667 enterococcu
5	70	12.7	307	12	Q9WQ79 felina infe
6	70	12.6	307	12	Q9WQ80 canine coro
7	69.5	12.6	532	16	Q8YD38 bruceella me
8	69.5	12.6	1792	13	Q57484 gallus galli
9	69	12.5	668	6	Q8WNU6 canis fami
10	69	12.5	1280	6	Q46605 canis fami
11	68	12.3	712	11	Q8VHD9 caenorhabdi
12	67.5	12.3	184	5	Q09943 caenorhabdi
13	67	12.2	307	12	Q9WQ75 canis fami
14	67	12.2	653	12	Q91BR0 spodoptera
15	67	12.2	1486	4	Q14637 homo sapien
16	67	12.2	2121	12	Q9YMB7 transmissib

17	67	12.2	6685	12	Q91W06 transmissib
18	66.5	12.1	81	10	Q24225 oryza sativ
19	66.5	12.1	251	16	Q92B05 streptomyc
20	66.5	12.1	342	4	Q96BE4 homo sapien
21	66.5	12.1	461	16	Q8ZDL1 yeastsia pe
22	66.5	12.1	996	3	Q8X001 neurospora
23	66	12.0	78	10	Q8W4V6 beta vulgar
24	66	12.0	111	5	Q95259 chimpanzee
25	66	12.0	210	10	Q84918 felis silve
26	66	12.0	387	12	Q8Q9Z0 schistosac
27	66	12.0	534	3	Q74996 caenorhabdi
28	66	12.0	1163	6	Q9TSU2 ralsontia s
29	65.5	11.9	197	3	Q9P7F1 mus musculi
30	65.5	11.9	233	5	Q45533 dictyosteli
31	65.5	11.9	382	16	Q8XT05 caenorhabdi
32	65.5	11.9	580	5	Q8U1U8 caenorhabdi
33	65.5	11.9	2346	11	Q9JUL1 mus musculi
34	65	11.8	199	5	Q94476 oryza sativ
35	65	11.8	256	10	Q8S290 oryza sativ
36	65	11.8	766	10	Q8W3Y6 persea amer
37	65	11.8	1086	4	Q75808 homo sapien
38	64.5	11.7	147	4	Q9HSU7 homo sapien
39	64.5	11.7	1007	13	Q902N3 gallus galli
40	64	11.6	54	5	Q9BK84 mesobuthus
41	64	11.6	225	10	Q8S4P7 tritricum ae
42	64	11.6	767	13	Q9DGR2 xenopus lae
43	64	11.6	1272	11	Q9JKE4 rattus norv
44	64	11.6	1272	11	Q92416 rattus norv
45	64	11.6	1551	16	Q8YMU3 anabaena sp

## ALIGNMENTS

## RESULT 1

Q05838 ID Q05838 PRELIMINARY: PRT; 486 AA.  
AC 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hexokinase PI (HKX2).  
GN HKX2 OR YGL253W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86083199; PubMed=3908224;  
RA Kopetzki E., Entian K.D., Mecke D.;  
RT "Complete nucleotide sequence of the hexokinase PI gene (HKX1) of  
RT Saccharomyces cerevisiae.";  
RL Gene 39:95-101(1985).  
DR EMBL, M14411; AAA34699.1; ..  
DR HSSP; P19367; IOHA.  
DR SGD; S0003222; HKX2.  
DR InterPro; IPR001312; Hexokinase.  
DR Pfam; PF00349; hexokinase2.1.  
DR Pfam; PF03727; hexokinase2.1.  
DR PRINTS; PR00475; HEXOKINASE.  
DR ProDom; PD001109; Hexokinase; 1.  
DR PROSITE; PS00378; HEXOKINASES; 1.  
SQ SEQUENCE 486 AA; 53952 MW; 218FAC9B8C36C8 CRC64;

Query Match 14.1%; Score 77.5; DB 3; Length 486;  
Best Local Similarity 31.2%; Pred. No. 2;  
Matches 30; Conservative 8; Mismatches 27; Indels 31; Gaps 6;

QY 8 FSVITVM---LIMASEMVNGSAFTWSPGCNNRAERYSKCCSAIHOXGDFSYTG 64  
DB 369 FGINTVGERKILRLSLIGA-----RAARLSVCIAIICQKRGK---TG 412

```

QY 65 QTAALYNQAGCGVAHTRF-GSSARACNP-----FGW 95
DB 413 HIAA-----DGSVSTRYPGFKEKAANALKDIYGM 441

RESULT 2
Q9BLR5 PRELIMINARY; PRT; 1558 AA.
AC Q9BLR5; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 165.8 kDa protein.
GN L654.05.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Toato V., Ciarlioni L., Blanchettin G., Bruschi C.V., Ivens A.C.,
RA Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL512294; CAC29453.1; -
DR InterPro; IPR001865; Ribosomal S2.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 1558 AA; 165803 MW; BB029575FDED8265 CRC64;

Query Match 13.7%; Score 75.5; DB 5; Length 1558;
Best Local Similarity 24.6%; Pred. No. 11;
Matches 32; Conservative 10; Mismatches 33; Indels 55; Gaps 6;

QY 10 VITVMTLIMASEMNGSAFTVWSGPGCNRAERY-SKCGCSAIHQKGYDFSFTGTAA 68
DB 663 VAALLSWAVDFEVRGLRYM-----RLAQRVHSRCPAKSCGEIGAD-----AAA 710
QY 69 LYN-----OAGCSGVAH--TRFGS-----S 86
DB 711 LHNSDPWRGVDAAVRCYTDAGCSGQAHGEGPALGSLGCLLLAIWLALVDVDDGVGA 770

QY 87 ARACNPFQWK 96
DB 771 AAPCPMTWK 780

RESULT 3
O45025 PRELIMINARY; PRT; 254 AA.
AC O45025;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Spinalin.
OS Hydra magnipapillata (Hydra).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Anthomedusae;
OC Hydroidea; Hydra.
OX NCBI_TaxID=6085;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF-1; TISSUE=NEMATOCYST SPINES;
RX MEDLINE=98248440; PubMed=9580562;
RA Koch A.W., Holstein T.W., Mala C., Kurz E., Engel J., David C.N.;
RT "Spinalin, a new glycine and histidine rich protein in spines of Hydra
nematocysts.";
```

```

RL J. Cell Sci. 111:1545-1554(1998).
DR EMBL; AF043907; AAC39121.1; -
SQ SEQUENCE 254 AA; 25393 MW; 649BAB00B3BEABF9 CRC64;

Query Match 13.6%; Score 75; DB 5; Length 254;
Best Local Similarity 27.3%; Pred. No. 1.9;
Matches 24; Conservative 14; Mismatches 24; Indels 26; Gaps 6;

QY 13 VMMLIAMASEMNGSAFTVWSGPGCNRAERY-SKCGCSAIHQKGYDFSFTGTAAALYNQ 72
DB 9 VLLLVAVDAR-----PW-GPGCAGSYGYGGCG---HHQANGY-----GGAHHA 48
QY 73 AG-CSGVAH-----TRFGSSARACNPFQ 94
DB 49 AGCCNGLAHGHHGGAYGOAAHHAGGYG 76

RESULT 4
O07667 PRELIMINARY; PRT; 730 AA.
AC O07667;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bp3s protein.
GN Bp3s.
OS Enterococcus hirae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SI85;
RX MEDLINE=99449055; PubMed=10520745;
RA Duez C., Thamm I., Sapunaric F., Coyette J., Ghuyesen J.M.;
RT "The division and cell wall gene cluster of Enterococcus hirae SI85.";
RL DNA Seq. 9:149-161(1998).
DR EMBL; Y13922; CAA74232.1; -
DR InterPro; IPR004325; PBP_Cterm.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF02968; PBP_C; 1.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
SQ SEQUENCE 730 AA; 80947 MW; C928FF1C10E810BB CRC64;

Query Match 12.9%; Score 71; DB 2; Length 730;
Best Local Similarity 46.8%; Pred. No. 16;
Matches 22; Conservative 4; Mismatches 13; Indels 8; Gaps 3;

QY 26 GSFTVWSGPGCNRAERY-SKCGCSAI-HQKGYDFSFTGTAAALYN 71
DB 520 GSAYGVTVSGYNVSA-----KTGTQIASDKGGYQ---TGDYALYS 559

RESULT 5
Q9WQ79 PRELIMINARY; PRT; 307 AA.
AC Q9WQ79;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RNA-directed RNA polymerase (Fragment).
GN POL.
OS feline infectious peritonitis virus (FIPV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11135;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCD2;
RX MEDLINE=99319897; PubMed=10392726;
RA Stephensen C.B., Casebolt D.B., Gangopadhyay N.N.;
```

RT "Phylogenetic analysis of a highly conserved region of the polymerase  
RT gene from 11 coronaviruses and development of a consensus polymerase  
chain reaction assay.";

RL Vireb Res. 60:181-189(1999).

DR EMBL: AF124987; AAD32991.1; -

KM RNA-directed RNA polymerase.

FT NON TER 307 307  
SQ SEQUENCE 307 AA; 34473 MW; 2A3AD4C643AC6E1 CRC64;

Query Match 12.7%; Score 70; DB 12; Length 307;  
Best Local Similarity 25.0%; Pred. No. 8.3;

Matches 26; Conservative 13; Mismatches 35; Indels 30; Gaps 4;

DB 16 L1AMASEMWSAFTWSSGCGNNRABRYSCG-----CSA1HQKGYDFSYTGOTAA 68

DB 124 MIRNASAMTIGSKHV-----GCCTHSDRFRSLNELAQLTEVHVCTGCFKFGGTTSG 178

QY 69 LYNQAGSGVAHTRF-----GSSARACNPFQWKS1 98

DB 179 ----DGTAVANSAFNIFQAVSANVKKLGVDSNACNNVTVKS1 218

RESULT 6

Q9WQ80 PRELIMINARY; PRT; 307 AA.

AC Q9WQ80; (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE RNA-directed RNA polymerase (Fragment).

OS POL.

OC Canine coronavirus.

CC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirinae.

OX NCBI\_TaxID=11153;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1-71;

RX MEDLINE=99319897; PubMed=10392726;

RA Stephenson C.B., Casbolt D.B., Gangopadhyay N.N.;

RT "Phylogenetic analysis of a highly conserved region of the polymerase  
RT gene from 11 coronaviruses and development of a consensus polymerase  
chain reaction assay.";

RL Vireb Res. 60:181-189(1999).

DR EMBL: AF124986; AAD32990.1; -

KM RNA-directed RNA polymerase.

FT NON TER 307 307  
SQ SEQUENCE 307 AA; 34500 MW; 2A3AD4C643AC9711 CRC64;

Query Match 12.7%; Score 70; DB 12; Length 307;  
Best Local Similarity 25.0%; Pred. No. 8.3;

Matches 26; Conservative 13; Mismatches 35; Indels 30; Gaps 4;

QY 16 L1AMASEMWSAFTWSSGCGNNRABRYSCG-----CSA1HQKGYDFSYTGOTAA 68

DB 124 MIRNASAMTIGSKHV-----GCCTHSDRFRSLNELAQLTEVHVCTGCFKFGGTTSG 178

QY 69 LYNQAGSGVAHTRF-----GSSARACNPFQWKS1 98

DB 179 ----DGTAVANSAFNIFQAVSANVKKLGVDSNACNNVTVKS1 218

RESULT 7

Q8YD38 PRELIMINARY; PRT; 532 AA.

AC Q8YD38; (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 21, Last annotation update)

DE Alcohol dehydrogenase (acceptor) (EC 1.1.99.-).

GN BME110339.

OS Brucella melitensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Brucellaceae; Brucella.

OX NCBI\_TaxID=29459;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=16W / ATCC 23456 / B10TYPE 1;

RX MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapattal V., Redkar R.J., Patra G., Mujar C., Los T.,

RA Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Reznik G.,

RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Coleman E.,

RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-D.,

RA Haselkorn R., Kyrides N., Overbeek R.,

RT "The genome sequence of the facultative intracellular pathogen  
Brucella melitensis.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).

DR EMBL: AE009671; AAL53581.1; -

DR InterPro: IPR00172; GNC\_OXRED.

DR Pfam: PF00732; GNC\_OXRED\_1.

DR PROSITE: PS00624; GNC\_OXRED\_2; 1.

KM Oxidoreductase; Complete proteome.

SQ SEQUENCE 532 AA; 58225 MW; F15B57645D149ADD CRC64;

Query Match 12.6%; Score 69.5; DB 16; Length 532;  
Best Local Similarity 27.2%; Pred. No. 17;

Matches 22; Conservative 8; Mismatches 16; Indels 35; Gaps 4;

QY 29 FTWSSGCGNNRABRYSC-----CGCSA1H-----QKGYDFSYTGOTALYNQAGSG 77

DB 64 FTTEAPGLNGRSLNYPKRGVLGGSSINGM1YMGQACDVG-----LMRQAGCGD 114

QY 78 VAHTRFGSSARACNPFQWKS1 98

DB 115 -----WGMDDV 120

RESULT 8

ID 057484 PRELIMINARY; PRT; 1792 AA.

AC 057484;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Laminin beta 2-like chain.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93015947; PubMed=1400373;

RA O'Rear J.J.;

RT "A novel laminin B1 chain variant in avian eye.";

RL J. Biol. Chem. 267:20555-20557(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=98209634; PubMed=9550264;

RA Liu J., Swadlow S., Xie W., Brewton R.G., Mayne R.;

RT "Primary structure and expression of a chicken laminin beta chain:  
evidence for four beta chains in birds.";

RL Matrix Biol. 16:471-481(1998).

DR EMBL: AF038555; AAB92586.1; -

DR HSPD: P02468; IKLO.

DR InterPro: IPR02106; AACRNA\_ligaseII.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR002049; Laminin\_EGF.

DR InterPro: IPR001886; LamNT.

DR Pfam: PF00053; Laminin\_EGF; 13.

DR Pfam: PF00055; Laminin\_Nterm; 1.

DR PRINTS: PR00011; EGF\_LAMININ.

DR SMART: SM00180; EGF\_Lam; 13.

DR SMART: SM00136; LamNT; 1.

```
DR PROSITE; PS00339; AA TRNA LIGASE II 2; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ TYPE_EGF; 12.
DR EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.
SQ SEQUENCE 1792 AA; 195723 MW; 4A4CBE0206F6A600 CRC64;

Query Match 12.6%; Score 69.5; DB 13; Length 1792;
Best Local Similarity 25.0%; Pred. No. 61;
Matches 25; Conservative 9; Mismatches 41; Indels 25; Gaps 4;

QY 18 AMASEMVNGSAFTVWSPGNC-NRAERY-----SKGCSAIHOKGGYD 59
DQ 300 ATAGVMVHGRCVCKHHTQGLNCERCDFVHELPWRRPAGSSNACRDCNHSRCHFD 359
DQ 60 ---FSTVGTQTAALYNOAGSGVAHTFGSSARACNPFPGWK 96
DQ 360 MAVFLATGNTSG----AVDCGQCHMTGRHCHLCKPFFYK 395

RESULT 9
Q8WNU6 PRELIMINARY; PRT; 668 AA.
AC Q8WNU6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Multidrug resistance p-glycoprotein (fragment).
GN MDRI.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 12.5%; Score 69; DB 6; Length 1280;
Best Local Similarity 25.5%; Pred. No. 48;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFFSVITVWMLIAMASEMV-----NGSAFTVWSPGNCNNRAERYSKGCSAIHOKGGYD 59
DQ 335 VFFSVLIGAFSGQASPIEAFANARGAAVEIFKIIDNKPSIDSYSGHKPDNIKGNLE 394
DQ 60 -----FSY-----TGQTAALYNOAGC 75
DQ 395 FKNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSGC 432

RESULT 11
Q8VHD9 PRELIMINARY; PRT; 712 AA.
AC Q8VHD9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Diversin.
GN ANKRD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 12.3%; Score 68; DB 11; Length 712;
Best Local Similarity 23.5%; Pred. No. 34;
Matches 24; Conservative 21; Mismatches 37; Indels 20; Gaps 3;

QY 1 MASTKLPFSVITVWMLIAMASEMVNGSAFTVWSPGNCNNRAERYSKGCSAIHOKGGYD 59
DQ 48 LAANKGHLSVQILLKAGCDLVDQDQDTALHRTAVVGNTEILTALIREGC--ALDRQD 105
```

```
DR PROSITE; PS00339; AA TRNA LIGASE II 2; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ TYPE_EGF; 12.
DR EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.
SQ SEQUENCE 1792 AA; 195723 MW; 4A4CBE0206F6A600 CRC64;

Query Match 12.6%; Score 69.5; DB 13; Length 1792;
Best Local Similarity 25.0%; Pred. No. 61;
Matches 25; Conservative 9; Mismatches 41; Indels 25; Gaps 4;

QY 18 AMASEMVNGSAFTVWSPGNC-NRAERY-----SKGCSAIHOKGGYD 59
DQ 300 ATAGVMVHGRCVCKHHTQGLNCERCDFVHELPWRRPAGSSNACRDCNHSRCHFD 359
DQ 60 ---FSTVGTQTAALYNOAGSGVAHTFGSSARACNPFPGWK 96
DQ 360 MAVFLATGNTSG----AVDCGQCHMTGRHCHLCKPFFYK 395

RESULT 9
Q8WNU6 PRELIMINARY; PRT; 668 AA.
AC Q8WNU6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Multidrug resistance p-glycoprotein (fragment).
GN MDRI.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 12.5%; Score 69; DB 6; Length 1280;
Best Local Similarity 25.5%; Pred. No. 48;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFFSVITVWMLIAMASEMV-----NGSAFTVWSPGNCNNRAERYSKGCSAIHOKGGYD 59
DQ 335 VFFSVLIGAFSGQASPIEAFANARGAAVEIFKIIDNKPSIDSYSGHKPDNIKGNLE 394
DQ 60 -----FSY-----TGQTAALYNOAGC 75
DQ 395 FKNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSGC 432

RESULT 11
Q8VHD9 PRELIMINARY; PRT; 712 AA.
AC Q8VHD9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Diversin.
GN ANKRD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 12.3%; Score 68; DB 11; Length 712;
Best Local Similarity 23.5%; Pred. No. 34;
Matches 24; Conservative 21; Mismatches 37; Indels 20; Gaps 3;

QY 1 MASTKLPFSVITVWMLIAMASEMVNGSAFTVWSPGNCNNRAERYSKGCSAIHOKGGYD 59
DQ 48 LAANKGHLSVQILLKAGCDLVDQDQDTALHRTAVVGNTEILTALIREGC--ALDRQD 105
```



QY 46 KCGSAIHOKGKYDFTGTOTALYNQACSGVAHTRGSSA 87  
 DB 106 KOGNTALHEAMHGFSS---GSAKLKAGANVLAARKKAGNTA 144

## RESULT 12

ID 009943 PRELIMINARY; PRT; 184 AA.

AC 009943;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE F12A10.1 protein.  
 GN F12A10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;  
 OC Rhabdilitidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;

SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ahnecough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,  
 Gardner A., Green P., Hawkins T., Hillier L., Jier A., Johnston L.,  
 Jones M., Kershaw J., Kiraten J., Laister N., Latreille P.,  
 Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Rifken L., Roopra A., Saunders B., Showkhen R.,  
 Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,  
 Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL Nature 368:32-38 (1994).

SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;  
 RA Gelsel C.;  
 RL "The sequence of C. elegans cosmid F12A10."  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RC STRAIN=BRISTOL N2;  
 RA Waterson R.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 SQ SEQUENCE 184 AA; 20062 MW; EF19A9A6C96B175A CRC64;

Query Match 12.3%; Score 67.5; DB 5; Length 184;  
 Best Local Similarity 24.1%; Pred. No. 9.3;  
 Matches 26; Conservative 11; Mismatches 34; Indels 37; Gaps 4;

QY 1 MASTKLFESYITWMLJLMASEVNV-----GSAFTWVG-----PG 36  
 DB 78 MASTLTFVFGILVAIVSLAFDLDLNPKNLQWRSSSGSVKRGCGGPGGCGG 137  
 QY 37 CNNAERYSKCGSAIHOKGKYDFTGTOTALYNQACSGVAHTRFG 84  
 DB 138 -----GSGYGV-----GCTGWTGCGMGSGSYSSWSYSSRSG 172

## RESULT 13

ID 09W075 PRELIMINARY; PRT; 307 AA.

AC 09W075;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RNA-directed RNA polymerase (Fragment).  
 GN POL.  
 OS Transmissible gastroenteritis virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11149;

SEQUENCE FROM N.A.

RX MEDLINE=99319897; PubMed=10392726;  
 RA Stephens C.B., Caebeolt D.B., Gangopadhyay N.N.;  
 RT "Phylogenetic analysis of a highly conserved region of the polymerase  
 RT gene from 11 coronaviruses and development of a consensus polymerase  
 RT chain reaction assay."  
 RL Virus Res. 60:181-189 (1999).  
 DR EMBL, AF124992; AAD32996.1;  
 KM RNA-directed RNA polymerase.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 307 AA; 34500 MW; B3215EE59E9A21EF CRC64;

Query Match 12.2%; Score 67; DB 12; Length 307;  
 Best Local Similarity 25.0%; Pred. No. 18;  
 Matches 26; Conservative 12; Mismatches 36; Indels 30; Gaps 4;

QY 16 LIAMSEWNGSAFTWVGSGCNRNARYKCG-----CSAIIHOKGKYDFTGTOTAA 68

DB 124 MIRNASAMILSSKXV-----GCTTHDRFYRLSNELAQVLTVEVHGTGFGYFGGTTG 178

QY 69 LYNQACSGVAHTRF-----GSSARACNPFEGMKSI 98  
 DB 179 -----GGTAIVNSAFNIFQAVSANVKNLGVDSNACNNVTKSI 218

## RESULT 14

ID 091BRO PRELIMINARY; PRT; 653 AA.

AC 091BRO;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE ORF131 p74.  
 OS Spodoptera exigua nucleopolyhedrovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=10454;

RP SEQUENCE OF 596-653 FROM N.A.  
 RX MEDLINE=93286555; PubMed=8509757;  
 RA Zuidema D., van Oers M.M., van Strien E.A., Caballero P.C., Klok E.J.,  
 RA Goldbach R.W., Vlak J.M.;  
 RT "Nucleotide sequence and transcriptional analysis of the p10 gene of  
 RT Spodoptera exigua nuclear polyhedrosis virus."  
 RL J. Gen. Virol. 74:1017-1024 (1993).

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20036646; PubMed=10567663;  
 RA Ikel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,  
 RA Goldbach R.W., Vlak J.M.;  
 RT "Sequence and organization of the spodoptera exigua multicapsid  
 RT nucleopolyhedrovirus genome."  
 RL J. Gen. Virol. 80:3289-3304 (1999).

RP SEQUENCE FROM N.A.  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
 RA Ikel W.F., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,  
 RA Goldbach R.W., Vlak J.M.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF169823; AAF33660.1;  
 SQ SEQUENCE 653 AA; 74213 MW; DADAD68F070FE0BC CRC64;

Query Match 12.2%; Score 67; DB 12; Length 653;  
 Best Local Similarity 33.3%; Pred. No. 40;  
 Matches 19; Conservative 9; Mismatches 15; Indels 14; Gaps 3;

QY 43 RYKCGCSAI-----HOKGYD-----FSYTGQTAALYNQAGC---SGVAHTRFGS 85  
DB 65 KFSKRGCSMTCPFFHETGPIDANTPANTQTSFTAILYAPACYNLDRVAATREGA 121

RESULT 15

014637 PRELIMINARY; PRT; 1486 AA.  
AC 014637;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Laminin alpha 3b chain (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98055454; PubMed=9395076;  
RA Doliana R., Bellina I., Buccioti F., Mongiat M., Perris R.,  
RA Colombatti A.;  
RT "The human alpha3b is a 'full-sized' laminin chain variant with a more  
RT widespread tissue expression than the truncated alpha3a.";  
RL FEBS Lett. 417:65-70(1997).  
DR EMBL; AF005258; AAC51867.1; .  
DR HSSP; P02468; 1TLE.  
DR InterPro; IPR000561; EGF-like  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR001886; LamNT.  
DR Pfam; PF00053; laminin EGF; 10.  
DR Pfam; PF00055; laminin Nterm; 1.  
DR PRINTS; PR00011; EGFLAMININ.  
DR ProDom; PD002082; LamNT; 1.  
DR SMART; SM00180; EGF\_Lam; 10.  
DR SMART; SM00001; EGF\_Like; 1.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF 1; UNKNOWN 10.  
DR PROSITE; PS01248; LAMININ TYPE EGF; 9.  
KW Glycoprotein; Laminin EGF-like domain; Repeat.  
FT NON\_TER 1  
FT NON\_TER 1486  
SQ SEQUENCE 1486 AA; 162496 MW; F2E5EB7EB522E98C CRC64;

Query Match 12.2%; Score 67; DB 4; Length 1486;  
Best Local Similarity 28.1%; Pred. No. 95;  
Matches 27; Conservative 8; Mismatches 23; Indels 38; Gaps 7;

QY 34 GPGCNN-RAEYVS-----KCGCSATHQKGYDF---SYTGO-----TAAL 69  
DB 480 GPRCDTCRSGFSFPTCQACWCSAL---GSYQMPCCSVTGOCECRPGVTGQRCRCLSGA 536  
QY 70 YNAGCGSVAHTRFGSSARACNPFQ---WKSIFIOQ 102  
DB 537 YDFPHCOQ-----SSSACDPAGTINWNLGYCQC 564

Search completed: January 12, 2003, 09:38:22  
Job time : 84 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2003, 07:06:54 ; Search time 2665 Seconds

(without alignments)  
5383.746 Million cell updates/sec

Title: US-09-882-434A-2

Perfect score: 493

Sequence: 1 atcaagctcttgagctccat.....ccatgcatgatttacc 493

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

GenEmbl: 1: gb\_ba: 2: gb\_hg: 3: gb\_in: 4: gb\_cm: 5: gb\_ov: 6: gb\_pac: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: gb\_vl: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_ov: 22: em\_pac: 23: em\_ph: 24: em\_pl: 25: em\_ro: 26: em\_sts: 27: em\_un: 28: em\_vl: 29: em\_vl: 30: em\_hg\_hum: 31: em\_hg\_in: 32: em\_hg\_inv: 33: em\_hg\_mus: 34: em\_hg\_pln: 35: em\_hg\_rod: 36: em\_hg\_mam: 37: em\_hg\_vtc: 38: em\_sy: 39: em\_hgo\_hum: 40: em\_hgo\_mus: 41: em\_hgo\_other:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	492.6	99.9	493	8	MIRNAPI
2	41.2	8.4	100919	9	AL359378
3	39.4	8.0	168575	9	AC091989
4	39.2	8.0	110913	2	CNS07YQ2
5	39.2	8.0	137936	2	CNS08CAX
6	39.2	8.0	144778	2	CNS07EPR
7	39	7.9	152137	2	AP001932
8	39	7.9	153485	2	AC090757
9	39	7.9	175032	9	AC019240
10	38.6	7.8	99862	2	AC106935
11	38.6	7.8	149767	2	AC113238
12	38.4	7.8	105470	2	AC116306
13	38.4	7.8	159919	9	AL359194
14	38.2	7.7	80828	5	AL672211
15	38.2	7.7	126118	9	AC104598
16	38.2	7.7	222454	2	AC098159
17	38.2	7.7	248360	2	AC099351
18	38	7.7	120832	2	CNS08C9W
19	38	7.7	128417	2	CNS08C4
20	38	7.7	167473	2	AC118111
21	37.8	7.7	140929	2	AC117805
22	37.2	7.5	224	8	AY024154
23	37.2	7.5	443	9	HSP26C11
24	37.2	7.5	29855	8	AP003826
25	37.2	7.5	192833	2	AL672160
26	37.2	7.5	204104	2	AL732613
27	36.8	7.5	114964	10	AL691507
28	36.8	7.5	177032	9	CNS06C7W
29	36.8	7.5	213564	2	AC095604
30	36.8	7.5	326980	10	AL591478
31	36.8	7.5	304322	2	AC118787
32	36.6	7.4	125020	9	AF429315
33	36.4	7.4	385	11	HS024XB8
34	36.4	7.4	76245	2	AC125151
35	36.4	7.4	110000	2	AC125102
36	36.4	7.4	150655	9	AC025887
37	36.4	7.4	157979	9	AC009835
38	36.4	7.4	166408	2	AC102255
39	36.4	7.4	167941	9	AC092638
40	36.4	7.4	170082	2	AC129774
41	36.4	7.4	195296	2	AC124448
42	36.4	7.4	266229	2	AC073676
43	36.2	7.3	68147	2	AC123627
44	36.2	7.3	69201	9	AC004109
45	35.2	7.3	85907	8	ATF28M11

## ALIGNMENTS

RESULT 1  
LOCUS MIRNAPI  
DEFINITION M.integrifolia mRNA for antimicrobial protein 1.  
ACCESSION Y10903  
VERSION Y10903.1 GI:2181942  
KEYWORDS  
SOURCE AMPI gene.  
ORGANISM Macadamia integrifolia.  
Macadamia integrifolia.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Proteaceae;  
Macadamia.  
REFERENCE 1 (bases 1 to 493)  
Marcus,J.P., Goulter,K.C., Green,J.L., Harrison,S.J. and  
Manners,J.M.

Prod. No. is the number of results predicted by chance to have a

TITLE Purification, characterisation and cDNA cloning of an antimicrobial peptide from *Macadamia integrifolia*

JOURNAL Eur. J. Biochem. 244 (3), 743-749 (1997)

MEDLINE 97261828

PUBMED 9108242

REFERENCE 2 (bases 1 to 493)

AUTHORS Marcus, J.P.

DIRECT SUBMISSION

TITLE Submitted (30-JAN-1997) J.P. Marcus, Cooperative Research Centre for Trop., Plant pathology., 5th Level John Hines Building,, The University of Queensland,, Brisbane, QLD 4072, AUSTRALIA

JOURNAL Location/Qualifiers

FEATURES

1..493

source /organism="Macadamia integrifolia"

/db\_xref="taxon:60698"

/clone="pGM7\*2"

/dev\_stages="seed"

70..378

gene /gene="AMPI"

70..378

CDS /gene="AMPI"

/function="antimicrobial protein"

/codon\_start=1

/protein\_id="CAA71842.1"

/db\_xref="GI:2181943"

/db\_xref="SWISS-PROT:P80915"

/translation="MASPKLFVITVVMILIAMSEVMVNGSAFTVWSPGCGNNRAERY SKCGSAIHQGGYDFSTYGTGTAALYNQAGCGVAHTRFGSSARACNPFPGWKSFIFQC"

sig\_peptide 70..147

mat\_peptide 148..375

BASE COUNT 132 a 106 c 109 g 145 t 1 others

ORIGIN

Query Match 99.9%; Score 492.6; DB 8; Length 493;

Best Local Similarity 100.0%; Pred. No. 3.2e-121;

Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGTCTTTGAGTCTCAGTATACATCTCTTCTCCGCCACCATAGCATTACAGCTA 60

DB 1 ATTAAGTCTTTGAGTCTCAGTATACATCTCTTCTCCGCCACCATAGCATTACAGCTA 60

QY 61 ACCTGAGCCATGCTCCACCAAGTTGTTCTTCTCAGTCATTACTGTGATGCTCAT 120

DB 61 ACCTGAGCCATGCTCCACCAAGTTGTTCTTCTCAGTCATTACTGTGATGCTCAT 120

QY 121 GCAATGCAAGTGCAGATGGTGAATGGAGTGCAATTTACAGATATGGAGTGGTCCAGTTGT 180

DB 121 GCAATGCAAGTGCAGATGGTGAATGGAGTGCAATTTACAGATATGGAGTGGTCCAGTTGT 180

QY 181 AACACCGTCTGAGCGATATAGCAAGTGGATGCTCAGCTATACATCAGAGGGAGGC 240

DB 181 AACACCGTCTGAGCGATATAGCAAGTGGATGCTCAGCTATACATCAGAGGGAGGC 240

QY 241 TATGACTTCAGCTACAGTGCACAACTGGCTCTCTACACAGGCTGGATGCTGAGTGT 300

DB 241 TATGACTTCAGCTACAGTGCACAACTGGCTCTCTACACAGGCTGGATGCTGAGTGT 300

QY 301 GTTGACACACACAGTTGGTCCAGTCCAGGCGATGCAACCCCTTTGGTGGAGAGT 360

DB 301 GTTGACACACACAGTTGGTCCAGTCCAGGCGATGCAACCCCTTTGGTGGAGAGT 360

QY 361 ATCTTCACCAAGTCTAGATTCATATCTTGGATCCATCTCTTATGTTTTCAGTGT 420

DB 361 ATCTTCACCAAGTCTAGATTCATATCTTGGATCCATCTCTTATGTTTTCAGTGT 420

QY 421 ATAATTAGAGAGTGCATGATATATAATAAAGTAAAGCTACGGTATCCCATGTG 480

DB 421 ATAATTAGAGAGTGCATGATATATAATAAAGTAAAGCTACGGTATCCCATGTG 480

QY 481 ATGATTTTACCC 493

DB 481 ATGATTTTACCC 493

## RESULT 2

AL359378

LOCUS

DEFINITION

Human DNA sequence from clone RP11-203H2 on chromosome 6 Contains

an STS and GSSs, complete sequence.

ACCESSION

AL359378

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

misc\_feature 4410..4909  
/note="match: GSS: Em:B82645"  
misc\_feature 4413..4740  
/note="match: GSS: Em:A0676785"  
repeat\_region 4697..5010  
/note="AluY repeat: matches 5..312 of consensus"  
repeat\_region 6390..6846  
/note="L1R3 repeat: matches 15..471 of consensus"  
repeat\_region 7809..7895  
/note="L2 repeat: matches 2559..2646 of consensus"  
repeat\_region 8184..8463  
/note="AluX repeat: matches 6..293 of consensus"  
repeat\_region 8755..8936  
/note="AluX repeat: matches 1..182 of consensus"  
repeat\_region 9424..9465  
/note="L21 copies 2 mer cc 78% conserved"  
repeat\_region 9467..9752  
/note="AluSg1 repeat: matches 1..300 of consensus"  
repeat\_region 9941..9970  
/note="L15 copies 2 mer ac 100% conserved"  
repeat\_region 9972..10031  
/note="L30 copies 2 mer at 83% conserved"  
repeat\_region 10704..10741  
/note="L2 repeat: matches 2713..2750 of consensus"  
repeat\_region 11433..12088  
/note="L1MC3 repeat: matches 7026..7693 of consensus"  
repeat\_region 12089..12549  
/note="MLT2FB repeat: matches 1..414 of consensus"  
repeat\_region 12550..12595  
/note="L1MC3 repeat: matches 7693..7737 of consensus"  
misc\_feature 12687..13148  
/note="match: GSS: Em:AQ529015"  
repeat\_region 15582..17530  
/note="L1M6 repeat: matches 4316..6297 of consensus"  
repeat\_region 17533..17765  
/note="L1M6 repeat: matches 4316..6297 of consensus"  
repeat\_region 17849..18864  
/note="L1M6 repeat: matches 3324..4338 of consensus"  
repeat\_region 18879..21256  
/note="L1P1 repeat: matches 3814..6123 of consensus"  
repeat\_region 21257..21564  
/note="AluY repeat: matches 5..307 of consensus"  
repeat\_region 21565..22057  
/note="L1P1 repeat: matches 3336..3814 of consensus"  
repeat\_region 22058..22367  
/note="AluY repeat: matches 1..311 of consensus"  
repeat\_region 22368..25902  
/note="L1P1 repeat: matches 1..3336 of consensus"  
repeat\_region 25902..26553  
/note="L1 repeat: matches 2673..3355 of consensus"  
repeat\_region 26679..26969  
/note="AluX repeat: matches 1..288 of consensus"  
repeat\_region 26993..27220  
/note="L1M2 repeat: matches 5031..5268 of consensus"  
repeat\_region 27267..27429  
/note="L1M2 repeat: matches 5031..5268 of consensus"  
repeat\_region 27431..27995  
/note="L1M2 repeat: matches 5284..5830 of consensus"  
repeat\_region 27996..28310  
/note="AluSg repeat: matches 1..313 of consensus"  
repeat\_region 28311..28547  
/note="L1M2 repeat: matches 5830..6070 of consensus"  
repeat\_region 30131..30479  
/note="L1M2 repeat: matches 10..365 of consensus"  
repeat\_region 31883..32342  
/note="L1R4 repeat: matches 1..490 of consensus"  
repeat\_region 32387..33129  
/note="L1R1 repeat: matches 2..785 of consensus"  
repeat\_region 33220..33532  
/note="AluY repeat: matches 1..296 of consensus"  
repeat\_region 33533..33573  
/note="L1R4 repeat: matches 479..519 of consensus"  
repeat\_region 35548..35844

/note="AluSg repeat: matches 1..297 of consensus"  
repeat\_region 36453..36763  
/note="AluY repeat: matches 1..310 of consensus"  
repeat\_region 36955..37295  
/note="L1P8 repeat: matches 5822..6162 of consensus"  
repeat\_region 37366..37399  
/note="L17 copies 2 mer gt 88% conserved"  
repeat\_region 38130..38308  
/note="MER5A repeat: matches 2..189 of consensus"  
repeat\_region 38709..39151  
/note="L1P10 repeat: matches 5719..6158 of consensus"  
misc\_feature 41095..41938  
/note="match: GSS: Em:AQ374227"  
/note="match: GSS: Em:AQ374227"  
/note="match: GSS: Em:AQ374468"  
repeat\_region 42099..42212  
/note="L57 copies 2 mer at 61% conserved"  
repeat\_region 42631..42935  
/note="AluY repeat: matches 1..310 of consensus"  
misc\_feature 43832..44317  
/note="match: GSS: Em:AQ564022"  
misc\_feature 43835..44056  
/note="match: GSS: Em:AQ282976"  
repeat\_region 44024..44534  
/note="L1M5 repeat: matches 5576..6152 of consensus"  
repeat\_region 45857..46065  
/note="MER3 repeat: matches 1..209 of consensus"  
repeat\_region 46372..46675  
/note="AluSg repeat: matches 1..294 of consensus"  
repeat\_region 46682..47334  
/note="L2 repeat: matches 1197..1913 of consensus"  
repeat\_region 47335..47440  
/note="MER81 repeat: matches 1..108 of consensus"  
misc\_feature 48086..48137  
/note="match: GSS: Em:AQ589537"  
repeat\_region 48974..49516  
/note="L26 copies 2 mer ac 96% conserved"  
repeat\_region 50479..50545  
/note="L2 repeat: matches 1474..2083 of consensus"  
repeat\_region 51008..51118  
/note="L1R repeat: matches 196..262 of consensus"  
repeat\_region 51087..51345  
/note="L1R repeat: matches 53..165 of consensus"  
repeat\_region 51187..51345  
/note="L1P16 repeat: matches 5913..6065 of consensus"  
misc\_feature 51483..51587  
/note="match: GSS: Em:AQ394900"  
repeat\_region 51634..51930  
/note="L1R repeat: matches 150..261 of consensus"  
repeat\_region 52522..52963  
/note="AluX repeat: matches 1..295 of consensus"  
misc\_feature 53173..53610  
/note="match: GSS: Em:AQ044785"  
repeat\_region 53193..53499  
/note="match: GSS: Em:AQ17553"  
repeat\_region 53575..53969  
/note="L2 repeat: matches 2267..2613 of consensus"  
misc\_feature 54881..55384  
/note="match: GSS: Em:AQ010823"  
/note="match: GSS: Em:AQ414205"  
misc\_feature 55495..55495  
complement(54955..55495)

Query March 8.4%; Score 41.2; DB 9; Length 100919;  
Best Local Similarity 50.5%; Pred.No.1.7; 94; Indels 0; Gaps 0;  
Matches 97; Conservative 1; Mismatches

QY 301 GTTGACACACAGGTTGGTCCAGTCCAGGACGATGACCTTTGGTTGGAAGT 360  
DB 42067 GTTGACATCTTTGGTTGTTACGATGCGTGTGACGATATATTTATGT 42126  
QY 361 ATTTTCATCCATGCTGATTTCAATCTTGGATCCATCTTTATGTTTCAAGT 420



Db 86357 AATAATCATGACTAAGTGAACCGTAATTAATTATAGTACTGTAGTAATTAAGAATATCT 86416  
OY 437 ATGATAT 444  
Db 86417 ATGGGTGT 86424

RESULT 5  
CNS08CAX 137936 bp DNA linear HTG 27-JUL-2002  
LOCUS Oryza sativa chromosome 12 clone OSUNBA0041K23, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AL844874  
VERSION AL844874.1 GI:22003300  
KEYWORDS HTG; HTGS PHASE2; HTGS\_ACTIVEFIN.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriacoidae; Oryzaceae; Oryza.  
1 (bases 1 to 137936)  
Choisne, N., Orjeda, G., Catolico, L., Demange, N., Wincker, P.,  
Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,  
Weissenbach, J. and Quetier, F.  
Oryza sativa chromosome 12 sequencing  
Unpublished  
2 (bases 1 to 137936)  
Genoscope.  
Direct Submission  
Submitted (26-JUL-2002) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
Web : www.genoscope.cns.fr)  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence.  
Work on the sequence is in progress and the release of this data is  
based on the understanding that the sequence may change as work  
continue. The sequence may be contaminated with foreign sequence  
from E.coli, yeast, vector, phage, etc.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1..137936  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="12"  
/clone="OSUNBA0041K23"  
/clone\_1ib="OSUNBA"  
BASE COUNT 38960 a 29563 c 29308 g 40105 t  
ORIGIN

Query Match 8.0%; Score 39.2; DB 2; Length 137936;  
Best Local Similarity 50.5%; Pred. No. 5.6;  
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

OY 257 CTGACAACTGCTGCTCTCTACCAACGAGCTGATGAGTGTGACACACAGGT 316  
Db 119273 CTAACTACTAATCTCTACATATCTGTCAACCAATTGCGCTCCAAATTCCTAAT 119214  
OY 317 TTGGGTCCAGTCCAGGCGATGCAACCCCTTTGGTGAAGATATCTTCATCAATGCT 376  
Db 119213 TCGCTTTCTTGTGTAAGAACTGCAAGCTGCGCAGAAAAGAGATGATTAATTGA 119154  
OY 377 AGATTCAATCTTGATTCATCTTCTATAGTTTTCAGAGTATTAATTGAAGATGC 436  
Db 119153 AAAATACATGACTAAGTGAACCGTAATTAATTAGTACTGTAGTAATTAAGAATATCT 119094  
OY 437 ATGATAT 444

Db 119093 ATGGGTGT 119086

RESULT 6  
CNS07EPR/c 144778 bp DNA linear HTG 06-FEB-2001  
LOCUS Oryza sativa chromosome 12 clone OSUNBA0041K23, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, in unordered pieces.  
ACCESSION AL513404  
VERSION AL513404.1 GI:12711302  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriacoidae; Oryzaceae; Oryza.  
1 (bases 1 to 144778)  
Salse, J., Choisne, N., Orjeda, G., Regad, F., Lorieux, M., Cooke, R.,  
Delserny, M., Robert, C., Broctier, P., Wincker, P., Cruaud, C.,  
Artiguenave, F., Saurin, W., Salanoubat, M., Quetier, F. and  
Weissenbach, J.  
Oryza sativa chromosome 12 sequencing  
Unpublished  
2 (bases 1 to 144778)  
Genoscope.  
Direct Submission  
Submitted (06-FEB-2001) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
Web : www.genoscope.cns.fr)  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence.  
Work on the sequence is in progress and the release of this data is  
based on the understanding that the sequence may change as work  
continue. The sequence may be contaminated with foreign sequence  
from E.coli, yeast, vector, phage, etc.  
Contigs composition :  
1062 bp contig from 1 to 1062  
933 bp contig from 2063 to 2995  
583 bp contig from 3996 to 4578  
2449 bp contig from 5579 to 8027  
66895 bp contig from 9028 to 75922  
67856 bp contig from 76923 to 144778.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
source  
1..144778  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="12"  
/clone="OSUNBA0041K23"  
/clone\_1ib="CUGi Nipponbare BAC"  
BASE COUNT 40362 a 30027 c 29723 g 39633 t 5033 others  
ORIGIN

Query Match 8.0%; Score 39.2; DB 2; Length 144778;  
Best Local Similarity 50.5%; Pred. No. 5.6;  
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

OY 257 CTGACAACTGCTGCTCTCTACCAACGAGCTGATGAGTGTGACACACAGGT 316  
Db 126114 CTAACTAGTAATCTCTACATATCTGTCAACCAATTGCGCTCCAAATTCCTAAT 126055  
OY 317 TTGGGTCCAGTCCAGGCGATGCAACCCCTTTGGTGAAGATATCTTCATCAATGCT 376  
Db 126054 TCGCTTTCTTGTGTAAGAACTGCAAGCTGCGCAGAAAAGAGATGATTAATTGA 125995  
OY 377 AGATTCAATCTTGATTCATCTTCTATAGTTTTCAGAGTATTAATTGAAGATGC 436  
Db 125994 AAAATACATGACTAAGTGAACCGTAATTAATTAGTACTGTAGTAATTAAGAATATCT 125935







```

Query Match      7.9%; Score 39; DB 9; Length 175032;
Best Local Similarity 63.2%; Pred. No. 6.2;
Matches 60; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 360 TATCTTTCACCAAGCTAGATTTTCATAACCTCTGGGATCCATCTTCTATCTTTTTTCAAGTG 419
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123352 TATATACATACAAGTATATATATACATATATACATGTATATATATATATATATATATATG 123411
QY 420 TATAATTAGAGAGATGCATGGATATATAATAAATA 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123412 TATATATACATATATACATGTATATATAATAATATA 123446

```

LOCUS AC106935 99862 bp DNA linear HTG 13-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-134K22, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 47 unordered pieces.  
 AC106935  
 VERSION AC106935.2 GI:21736698  
 KEYWORDS HTG; HTGS PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 99862)  
 REFERENCE  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Albrechts,S.L., Anaratunga,H.C., Are,D.R., Ayele,M., Banks,T.,  
 Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Koryan,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzner,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbar,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newson,N., Nguyen,A., Nguyen,S.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
 Orygunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., Rojiboken,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoshari,N., Sisson,I.,  
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vaequez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 99862)  
 Worley,K.C.  
 Direct Submission  
 Submitted (14-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 99862)  
 Worley,K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18141352.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information

Center project name: COMU  
 Center clone name: CH230-134K22  
 ----- Summary Statistics  
 Sequencing vector: Plasmid:  
 Chemistry: Dye-terminator Big Dye 100k of reads  
 Assembly program: Phrap; Version 0.990329  
 Consensus quality: 54814 bases at least Q40  
 Consensus quality: 59343 bases at least Q30  
 Consensus quality: 63485 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 47 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1307: contig of 1307 bp in length  
 \* 1308 1407: gap of unknown length  
 \* 1408 2662: contig of 1255 bp in length  
 \* 2663 2763: gap of unknown length  
 \* 2763 4238: contig of 1476 bp in length  
 \* 4239 4339: gap of unknown length  
 \* 4339 5561: contig of 1223 bp in length  
 \* 5562 6808: gap of unknown length  
 \* 6808 6908: contig of 1147 bp in length  
 \* 6908 8196: gap of unknown length  
 \* 8196 8297: contig of 1288 bp in length  
 \* 8297 9357: gap of unknown length  
 \* 9357 9457: contig of 1060 bp in length  
 \* 9457 10642: gap of unknown length  
 \* 10642 10741: contig of 1185 bp in length  
 \* 10741 11824: gap of unknown length  
 \* 11824 11925: contig of 1083 bp in length  
 \* 11925 13038: gap of unknown length  
 \* 13038 13137: contig of 1113 bp in length  
 \* 13137 14143: gap of unknown length  
 \* 14143 14242: contig of 1005 bp in length  
 \* 14242 15760: gap of unknown length  
 \* 15760 15761: contig of 1518 bp in length  
 \* 15761 15860: gap of unknown length  
 \* 15860 17577: contig of 1717 bp in length  
 \* 17577 17678: gap of unknown length  
 \* 17678 19577: contig of 1900 bp in length  
 \* 19577 19677: gap of unknown length  
 \* 19677 20886: contig of 1209 bp in length  
 \* 20886 20987: gap of unknown length  
 \* 20987 22146: contig of 1160 bp in length  
 \* 22146 22246: gap of unknown length  
 \* 22246 24293: gap of unknown length  
 \* 24293 24393: contig of 2047 bp in length  
 \* 24393 25431: gap of unknown length  
 \* 25431 25531: contig of 1038 bp in length  
 \* 25531 27240: gap of unknown length  
 \* 27240 27340: contig of 1709 bp in length  
 \* 27340 28620: gap of unknown length  
 \* 28620 28720: contig of 1280 bp in length  
 \* 28720 28721: gap of unknown length  
 \* 28721 30172: gap of unknown length  
 \* 30172 30272: contig of 1452 bp in length  
 \* 30272 31788: gap of unknown length  
 \* 31788 31888: contig of 1516 bp in length  
 \* 31888 33111: gap of unknown length  
 \* 33111 34312: contig of 2423 bp in length  
 \* 34312 34411: gap of unknown length  
 \* 34411 36437: contig of 2026 bp in length  
 \* 36437 36537: gap of unknown length  
 \* 36537 38524: contig of 1987 bp in length  
 \* 38524 38624: gap of unknown length  
 \* 38624 39719: contig of 1095 bp in length  
 \* 39719 39819: gap of unknown length  
 \* 39819 40934: contig of 1115 bp in length

```
* 40935 41034: gap of unknown length
* 41035 contig of 1838 bp in length
* 42872: contig of unknown length
* 42972: gap of unknown length
* 44956: contig of 1984 bp in length
* 44957: gap of unknown length
* 45056: gap of unknown length
* 48346: contig of 3290 bp in length
* 48347: gap of unknown length
* 48348: gap of unknown length
* 48447: contig of 2737 bp in length
* 51183: gap of unknown length
* 51283: gap of unknown length
* 53555: contig of 2272 bp in length
* 53556: gap of unknown length
* 53557: gap of unknown length
* 56150: contig of 2495 bp in length
* 56250: gap of unknown length
* 59131: contig of 2881 bp in length
* 59231: gap of unknown length
* 61673: contig of 2442 bp in length
* 61773: gap of unknown length
* 64165: contig of 2392 bp in length
* 64166: gap of unknown length
* 64265: gap of unknown length
* 66077: contig of 1812 bp in length
* 66177: gap of unknown length
* 68907: contig of 2730 bp in length
* 69007: gap of unknown length
* 72063: contig of 3056 bp in length
* 72163: gap of unknown length
* 74260: contig of 2097 bp in length
* 74360: gap of unknown length
* 78074: contig of 3714 bp in length
* 78174: gap of unknown length
* 80314: contig of 2140 bp in length
* 80414: gap of unknown length
* 83281: contig of 2867 bp in length
* 83381: gap of unknown length
* 87855: contig of 4404 bp in length
* 87885: gap of unknown length
* 91507: contig of 3622 bp in length
* 91607: gap of unknown length
* 94867: contig of 3260 bp in length
* 94967: gap of unknown length
* 94968 99862: contig of 4895 bp in length.

FEATURES             Location/Qualifiers
   source             1..99862
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-134K22"
BASE COUNT      27849 a 19699 c 19114 g 28488 t 4712 others
ORIGIN

Query Match      7.8%; Score 38.6; DB 2; Length 99862;
Best Local Similarity 32.0%; Pred. No. 8.3;
Matches 83; Conservative 1; Mismatches 175; Indels 0; Gaps 0;

QY 231 GAAGGAGGCTATGCTTACAGTACACTGGACAACTGCTCTCTACACAGGCTGG 290
Db 53529 GAGGCGGAGCAGAAAGAGGAGAGGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 53588

QY 291 ATGCACTGGTGTGCACACACACACAGGTTTGGGTCCAGTGCAGGGCATGCAACCCCTTTTGG 350
Db 53589 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 53648

QY 351 TTGGAAGAGTATCTTCATCAATGCTAGATTTTCATCAACTCTTGGATCCATCTTCATGTT 410
Db 53649 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 53708

QY 411 TTTCAGTGTATATTTAGAGAGTGCATGATATATAATAATAAGTAAAGTAAAGTAAAGTAA 470
Db 53709 TTTTATTCATATCTTATATACATATTTTCATATATATATATATATATATATATATATAT 53768

QY 471 TCACCATGTGATGATTTT 489
Db 53769 GCACATGGTCAAGCATGTT 53787
```

```
RESULT 11
AC113238
LOCUS
DEFINITION
AC113238
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
AC113238 149767 bp DNA linear HTG 27-FEB-2002
Felis catus clone RP86-252E8, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
AC113238
AC113238.1 GI:18958665
HTG; HTGS_PHASE1; HTGS_DRAFT.
Felis catus.
Felis catus
Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 149767)
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,
Masriani,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgonov,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 149767)
Green,E.D.
Direct Submission
Submitted (27-FEB-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hgri.nih.gov
----- Project Information
Center project name: cou
Center clone name: 252E08
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 147596 bases at least Q40
Consensus quality: 147747 bases at least Q30
Consensus quality: 147904 bases at least Q20
Insert size: 135000; agarose-fp
Quality coverage: 149367; sum-of-contigs
Quality coverage: 10.56x in Q20 bases; agarose-fp
Quality coverage: 9.55x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9118: contig of 9118 bp in length
* 9119 9218: gap of unknown length
* 9219 17046: contig of 7828 bp in length
* 17047 17146: gap of unknown length
* 17147 34899: contig of 17753 bp in length
* 34900 34999: gap of unknown length
* 35000 57007: contig of 22008 bp in length
* 57008 57107: gap of unknown length
* 57108 149767: contig of 92660 bp in length.
Location/Qualifiers
1..149767
/organism="Felis catus"
/db_xref="taxon:9685"
/clone="RP86-252E8"
/clone_lib="RP86"
misc_feature 1..9118
```



```
/db_xref="GI:19807760"
/translation="MIFDSLISLIPISMIMWGCCSNVLSLEIMKQSQSHALVTFFQ
FATVAFINRWKQVSIWIFIGLRRKIPKLTFLMVISIIFILSNLKKALDC
DIPPHMIFRSSLLSTVIGSIFYKSGYSQQILSLMVLTLGIFATFSSMPDCK
EISLGHENLRFSGIMLMAAPFLSSILGLIOETHYKLGKORHYETIFYSVNDLI
HHQLQNDLSALMALPGFGSPFTLWYLIVNVLVVCVTCGVFIITGKTSTLCTLVI
SIRKLSIIISYIFNNHTLSLFTGTLVFLGTFMYSTSGVKIEPLPTQVKEIE
KERKLN"
complement(23858...24619)
/notes="ORF ID:dd_00658"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99330.1"
/db_xref="GI:19807761"
/translation="MKILITFILLINLIIVKSLVVKTVNYSPGCTGGVKNVILHSH
CNLVFRELVMNVSYFANFKMDSCKNAKTNYAEPLHVGDEDNQIKVSYNYLTQY
DAINDLPMSDCNEVINGGCEENYIITSILKGTCSPIKYNDTEAMTYSYCDNSMLSY
ICSEPTFCIGNCAINPYKNSQCVPGRLDGNADGIYKFPVKLAKSQPKSPPTP
SQSTPISDEDEPNHKSVPASIALIIFGLLSL"
complement(join(25977...30925,31064...31175))
/notes="ORF ID:dd_00657"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99331.1"
/db_xref="GI:19807762"
/translation="MILRLNPKPIITFLFLLVLLNENVFSQQPYQYTCSEMEPIV
GDMGFSVIALSHPNISATSPWKATSVSFCTDDMGSSAYSLAGYIYQAFLYNESDPN
TQSTLTWTVSLPPLIPLEQYVLEYSOFETNINPVLNVDRPVASYNGNLLG
DLRVLPLESLPANIAPTTSYNNNGGGGGGGTQSGFRIGSVIIRKPAATWPEE
VYVDVAKSGKNGTSAVPSIIQQALSNDFNGTILYADCTYCNLSYLPISRSIN
IISQSGDPTKILSGEQLNLFKISYNSQDDFLTPYRRGRFRQMEYFIDISNLI
TSISGSGSGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
GGAALFEGNVASFNCIFTSNADQNRGSIARGNAALILENTLVQNCSCVASF
GPAISMLGFGSPITISLVNCTMENLIVNIDSTSPNMODNSIOGIEPIEGASLCS
MLRSTQSTNCDIGINNACILISSTSNITNFGNLSKPGNIGIYSQSSV
TIQNVLEDNNGNIGAIIPAMELSKVNIENTIIRNIGITNSLSSSDSLVLSQSIY
NTNDYTYLFFGSFKLFDNINSGKSYATQTVISISTLFDLISGSHVPODCQV
EITDSIFNLGFSFYSSKIHIVLSGSIENITQFLMISPLNSVDFYVYQVSDTII
INLLSQNLVAINSGKIDFNSTNSILQIYEGSTSTFKNLFQNRNPFGLGGCIF
TFSSVLDNCTAIGNFAKQGVSTIAGNLITIDGIIYONNSAGGGVIFYSTPG
PTIKNAIFSGNTASGNNTAADAALIDVEDPPSSFGSDQYSGRIYKQEGQIV
SGFVKLQIISGNTISTSTYNGIALFISNTLYGPIINTVDYFELNNPFLPVQTV
SNLKSFGQISPTIKTECTRGSYNGETCVQCPSHIAYCPGAKVNPQEGWHY
DVNQTPKPYLYECNPVIGSNTSCMPCAFSPGLCAQCDPGYDWMGCEKCDKTS
PILFIATVAGISGIMVLVQSDSSEGLMTVIIYFAQTLMLVSTGVKFNFLALLNLFQ
PGSSSIFGICPGPDYQHYFTLAPLWILVLLISTVFFIVYVRYKLIQKLNKI
IKQEDIEKSKSIISVSSIIIVTSDPGTTINSSILKSLERKLLRQKQOQOQPLEND
DDFSTNTTNDGDLRSDYSIADGGKHSILGKSIETISIGSLSQVTLIDNATAI
PSNQTFLSQQLGLMLNIIYPIGKATFELFCENVGSVATVILANGVSCYTDQY
HKALMVSCLLIVIGPLILMILLFLNRKSLDDPHQTYGVFIKLYKSSVYFDVI
LLRLRLIILMSTMDPTSAARSLVSVSVLLQKQYQFKRISDNRLSLTSLLL
FICCVLDNTVYQSIQEWIIVSSIIIFLHVVLYSNRYQIISLQHYLYIISRGK
KGRKSNWTLFGSNSLYNQDHYDDDFSRSGIFSKKQKSDTKTVKSKTNN
NENNIENNENNENNENNENNENNVKIEIKLIEESKTNFELNNYDDDDFDEDESS
DDEEDDNGESTNLLKP"
join(34837...35014,35188...35398,35512...36769)
/notes="ORF ID:dd_00653"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99332.1"
/db_xref="GI:19807763"
/translation="MSEIDTNDIKKEMDNKNRDPDPTDNKELLKWKTKKRKSNDSMHI
NNFYRNPNEFKLIANDKSNATKELTRVLDDHDFGLRIELPNLYCPTLTLRLNYL
WISQNLKTLILANDNDNKIIKIGIDIGTGTCTIFLLGKALFNWMSFGIDIDDKV
LEYAQNITINSLSKITLPNEKNSDILLKLANYSKSTNSNDHDDHDDDDDD
DEEPAFCPLPPFPKDLNENNNNNNNPKSCTGSNEMVTDGCEFEVKELIKES
FOLKCKRFTYTMIGRKNVNLNLIKQILPNQIQITTELVOGNTSRWLSWFL
NKSTNLETKENNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
IVGSSIKLDRDIEFLFTIFDITLRLILFKPIDPKINNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNSCDGEIINSLFFILKFLNLIKVEIKL"
37054...38214
/notes="ORF ID:dd_01574"
```

CDS

CDS

CDS

CDS

```
/codon_start=1
/product="PAP-1. 6/101"
/protein_id="AAL99333.1"
/db_xref="GI:19807764"
/translation="MDRRDYSKEKYKYGDNDKORRFSNDKIRHNFDEFTMPNDRDR
DKDIDRYDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR
YNSNSSSSPDDGKYKMDMRDRDKYREKRKFDSSRNNEHDRSSIRPLSYDGVDDDDY
GRSKNKTEQYKEMVASNGNKRERLEDSIPDLPENREAREPLRNAPTGLHQPLGK
EVKVMQCKKCFYHRAICTQTNPAEYIKREKLEKREKLEKREKLEKREKLEKREK
NNNNNNNNFTHKDRINKNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
EKKRKRKKEKKIKASKKSHRHSDDSDSGSGSSGSSDSDSLSKDESYSSNESE"
join(39790...40108,40252...47744)
/notes="ORF ID:dd_01572"
/codon_start=1
/product="Putative Acyltransferase"
/protein_id="AAL99334.1"
Query Match 7.8%; Score 38.4; DB 2; Length 105470;
Best Local Similarity 56.2%; Pred. No. 9.4;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 360 TATCTTCATCCATAGATTTCATAACTCTTGGATCCATCTTCTATGTGTTTCAAGTG 419
Db 93202 TATTTGAATCAATTTTAAATTTATATTGCTTTTATTCATAATGTTGATTAATAA 93261
QY 420 TATAATTAGAGAGATCGATGATATATAATAAGTAAAGCTACCGTATCACCATGT 479
Db 93262 AATAATTGAAAAAATAAGATAATAATTAATAATAATCTCAATATACACATGT 93321
QY 480 GATGATTT 487
Db 93322 AATTTTT 93329
RESULT 13
AL359194 Human DNA sequence from clone RP11-326D19 on chromosome 13,
LOCUS complete sequence.
AL359194
VERSION AL359194.7 GI:13273756
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159919)
AUTHORS Blakey, S.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Mar 11, 2001 this sequence version replaced gi:12963937.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
```

Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Ch13>  
RP11-326D19 is from the library RP11-11.2 constructed by the group  
of Pictet de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACE3.6  
This sequence is the entire insert of clone RP11-326D19. The true  
right end of clone RP11-22917 is at 61034 in this sequence.

## FEATURES

```
source
1..159919
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="13"
  /clone_id="RP11-326D19"
  /clone_lib="RP11-11.2"
2..1157
  /note="L1PA5 repeat: matches 4987. .6143 of consensus"
repeat_region
7130..7197
  /note="17 copies 4 mer atag 88% conserved"
repeat_region
7136..7195
  /note="30 copies 2 mer at 75% conserved"
repeat_region
7273..7356
  /note="21 copies 4 mer atata 71% conserved"
repeat_region
7276..7345
  /note="35 copies 2 mer ta 75% conserved"
repeat_region
7578..7641
  /note="L2 repeat: matches 2574. .2638 of consensus"
repeat_region
7769..7796
  /note="L2 repeat: matches 1323. .1542 of consensus"
repeat_region
8427..8466
  /note="L1PA5 repeat: matches 6035. .6074 of consensus"
repeat_region
9056..9128
  /note="L1MB3A repeat: matches 991. .1062 of consensus"
repeat_region
9847..10246
  /note="HAL1 repeat: matches 991. .1062 of consensus"
repeat_region
10247..10465
  /note="L1P repeat: matches 1863. .2261 of consensus"
repeat_region
11012..11318
  /note="L1PA5 repeat: matches 5920. .6138 of consensus"
repeat_region
12421..12694
  /note="ALuB repeat: matches 1. .298 of consensus"
repeat_region
13079..13536
  /note="L1P5 repeat: matches 5396. .5669 of consensus"
repeat_region
13537..13829
  /note="L1MD2 repeat: matches 5638. .6094 of consensus"
repeat_region
13830..13977
  /note="ALuS repeat: matches 5. .300 of consensus"
repeat_region
14104..14291
  /note="L1MD2 repeat: matches 5491. .5638 of consensus"
repeat_region
16400..16453
  /note="ALuB repeat: matches 93. .282 of consensus"
repeat_region
16425..16452
  /note="27 copies 2 mer ta 77% conserved"
repeat_region
16821..16856
  /note="7 copies 4 mer atata 100% conserved"
repeat_region
20660..21162
  /note="18 copies 2 mer tt 86% conserved"
repeat_region
21807..22019
  /note="L1MB8 repeat: matches 5649. .6161 of consensus"
repeat_region
22022..22105
  /note="L1 repeat: matches 4045. .4250 of consensus"
repeat_region
22202..22888
  /note="21 copies 4 mer gaag 77% conserved"
repeat_region
22889..23252
  /note="L1M4 repeat: matches 3360. .4073 of consensus"
repeat_region
23253..23302
  /note="L1 repeat: matches 1. .365 of consensus"
repeat_region
23303..23693
  /note="L1M4 repeat: matches 3310. .3360 of consensus"
repeat_region
23694..23996
  /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region
24000..24037
  /note="L1M4 repeat: matches 3074. .3310 of consensus"
  /note="19 copies 2 mer tt 81% conserved"
repeat_region

repeat_region
24040..24808
  /note="L1PA3 repeat: matches 5378. .6146 of consensus"
repeat_region
25498..26057
  /note="L1 repeat: matches 4. .541 of consensus"
repeat_region
26649..26903
  /note="ALuS repeat: matches 1. .255 of consensus"
repeat_region
27125..27152
  /note="7 copies 4 mer atag 92% conserved"
repeat_region
27432..27509
  /note="MADE1 repeat: matches 8. .80 of consensus"
repeat_region
28572..28675
  /note="L2 repeat: matches 210. .314 of consensus"
repeat_region
29376..31199
  /note="L1 repeat: matches 3165. .4993 of consensus"
repeat_region
31271..31678
  /note="L1 repeat: matches 1. .389 of consensus"
repeat_region
31702..31739
  /note="L19 copies 2 mer aa 78% conserved"
repeat_region
31773..31867
  /note="MER44 repeat: matches 1. .96 of consensus"
repeat_region
31871..32131
  /note="MER44 repeat: matches 474. .724 of consensus"
repeat_region
32405..32952
  /note="L1MB1 repeat: matches 5495. .6052 of consensus"
repeat_region
32954..33113
  /note="L2 repeat: matches 332. .510 of consensus"
repeat_region
35044..35456
  /note="L1 repeat: matches 332. .510 of consensus"
repeat_region
35578..35916
  /note="THE1C repeat: matches 7. .371 of consensus"
repeat_region
36028..36144
  /note="MIR repeat: matches 1. .120 of consensus"
repeat_region
36974..37277
  /note="ALuB repeat: matches 1. .297 of consensus"
repeat_region
37492..37591
  /note="L50 copies 2 mer gt 68% conserved"
repeat_region
37595..37800
  /note="L2 repeat: matches 2082. .2310 of consensus"
repeat_region
38474..38589
  /note="MIR repeat: matches 20. .141 of consensus"
repeat_region
39167..39460
  /note="ALuS repeat: matches 9. .312 of consensus"
repeat_region
41933..42581
  /note="MER67C repeat: matches 1. .709 of consensus"
repeat_region
43968..43995
  /note="7 copies 4 mer tctg 96% conserved"
repeat_region
44350..44381
  /note="16 copies 2 mer at 93% conserved"
repeat_region
44397..44493
  /note="HERV16 repeat: matches 2566. .2674 of consensus"
repeat_region
44656..44823
  /note="MER58A repeat: matches 51. .224 of consensus"
repeat_region
44979..45156
  /note="HERV1 repeat: matches 3020. .3198 of consensus"
repeat_region
45756..45854
  /note="L1 repeat: matches 3. .102 of consensus"
repeat_region
45888..46113
  /note="HERV1A0 repeat: matches 1124. .1351 of consensus"
repeat_region
46752..46979
  /note="HERV1A0 repeat: matches 1380. .1617 of consensus"
repeat_region
46979..49348
  /note="HERV1A0 repeat: matches 2050. .4496 of consensus"
repeat_region
49989..50162
  /note="HERV1A0 repeat: matches 5352. .5530 of consensus"
repeat_region
50166..50607
  /note="L1 repeat: matches 4. .458 of consensus"
repeat_region
50613..50815
  /note="L2 repeat: matches 2535. .2748 of consensus"
repeat_region
51701..51876
  /note="ALuS/x repeat: matches 129. .302 of consensus"
repeat_region
51893..52011
  /note="L1M4 repeat: matches 4110. .4231 of consensus"
repeat_region
52868..52899
```

```

/note="16 copies 2 mer gt 93% conserved"
54804..54928
/note="L2 repeat: matches 1714..1838 of consensus"
55708..55771
/note="16 copies 4 mer tatg 78% conserved"
58435..58781
/note="LTR33 repeat: matches 48..398 of consensus"
61425..62137
/note="L1MA9 repeat: matches 5549..6270 of consensus"
62138..62198
/note="L1MA9 repeat: matches 5317..5375 of consensus"
62235..62544
/note="AluY repeat: matches 1..308 of consensus"
62545..62669
/note="FLAM-A repeat: matches 2..131 of consensus"
62687..63166
/note="L1MA9 repeat: matches 4826..5309 of consensus"

Query Match      7.8%; Score 38.4; DB 9; Length 159919;
Best Local Similarity 55.4%; Pred. No. 9.1;
Matches 72; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

```

```

QY 360 TATCTTCATCCATGCTAGATTTCATATACTCTGTGATCCATCTCTCTATGTTTCAAGTG 419
DB 69212 TCTCTTAAGAGAGGTCCTCAATCTCTTAATCTCTGCAACTACTAATTTGTTTATTCATGTC 69271
QY 420 TATAATTTAGAGAGATGCGATGATATATAATTAATTAAGTAAAGCTACGGTATCACCATGT 479
DB 69272 TACAATTCGAGTATTAGAGAACATATATAATTTAATTTACATATAATGTAATGTAATGTTTGT 69331
QY 480 GATGATTTT 489
DB 69332 GATTTTCTTT 69341

```

```

RESULT 14
AL672211/C
LOCUS AL672211 80828 bp DNA linear VRT 16-AUG-2002
DEFINITION Zebrafish DNA sequence from clone BUSM1-6A2. Contains a novel gene similar to OGRF, a novel OGRF pseudogene, two pseudogenes on a novel LINE-like transposon, a novel reverse transcriptase pseudogene, the fez gene for forebrain embryonic zinc-finger protein, part of a novel gene for a protein similar to Ca2+ dependent activator and four CpG islands, complete sequence.
ACCESSION AL672211
VERSION AL672211.7 GI:20338538
KEYWORDS HTG; Ca2+ dependent activator; CpG island; fez; forebrain; OGRF; opioid growth factor receptor; transposon; zinc-finger protein.
SOURCE zebrafish
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 80828)
Lloyd.D.
Direct Submission
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Apr 29, 2002 this sequence version replaced gi:20268901.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following

```

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) BUSM1-6A2 is from a Zebrafish PAC library VECTOR: PCYPAC-6

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zface@sanger.ac.uk

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.Projects/D\\_rerio/fishmask.shtml](http://www.Projects/D_rerio/fishmask.shtml) his clone was isolated from a PAC library (Incyte Genomics, Inc.) and provided by G. Levkowitz (Genentech Inc., San Francisco, USA). This sequence is the entire insert of clone BUSM1-6A2. The true left end of clone BUSM1-113 is at 44958 in this sequence.

```

FEATURES
source
    1..80828
        /organism="Danio rerio"
        /db_xref="taxon:7955"
        /clone="BUSM1-6A2"
        /clone_lib="BUSM1"
        /note="2.0 copies 6 mer CATATA 24% conserved"
    129..153
        /note="2.8 copies 9 mer CATTATTAT 34% conserved"
    138..160
        /note="7.7 copies 3 mer ATT 28% conserved"
    458..477
        /note="2.0 copies 10 mer AGTTTTCAT 40% conserved"
    541..557
        /note="4.2 copies 4 mer AAT 34% conserved"
    553..566
        /note="2.0 copies 7 mer TACAAAA 28% conserved"
    587..709
        /note="1.9 copies 12 mer TACTTAAATAC 46% conserved"
    702..717
        /note="2.0 copies 8 mer TTAATAA 32% conserved"
    806..822
        /note="2.4 copies 7 mer ATAAAT 25% conserved"
    856..874
        /note="2.1 copies 9 mer TTGTGACTT 38% conserved"
    1084..1131
        /note="3.4 copies 14 mer TACACATACATACA 60% conserved"
    1088..1098
        /note="2.8 copies 4 mer CATG 22% conserved"
    1096..1107
        /note="2.0 copies 6 mer CATACA 24% conserved"
    1101..1133
        /note="8.2 copies 4 mer ACAT 66% conserved"
    1338..1405
        /note="Dr00340 repeat: matches 120..187 of consensus"
    1782..1799
        /note="18.0 copies 1 mer A 36% conserved"
    1815..1826
        /note="2.0 copies 6 mer GTCTCA 24% conserved"
    2405..2415
        /note="2.2 copies 5 mer AACAG 22% conserved"
    2552..2562
        /note="2.8 copies 4 mer TTGT 22% conserved"
    2787..2950
        /note="4.1 copies 40 mer ACACACTCATATACATATAAACTTGATGTCATGCAACAC 149% conserved"
    2841..2852
        /note="2.0 copies 6 mer CATATA 24% conserved"
    2935..2946
        /note="2.0 copies 6 mer CATATA 24% conserved"

```



repeat_region	/note="2.8 copies 9 mer CATTATTTAT 34% conserved"
repeat_region	2997..3019 /note="7.7 copies 3 mer ATT 28% conserved"
repeat_region	3317..3336 /note="2.0 copies 10 mer AGTTTACAT 40% conserved"
repeat_region	3402..3414 /note="3.2 copies 4 mer AAAT 26% conserved"
repeat_region	3510..3523 /note="2.0 copies 7 mer TACAAA 24% conserved"
repeat_region	3544..3566 /note="1.9 copies 12 mer TACTTAAATAC 46% conserved"
repeat_region	3559..3574 /note="2.0 copies 8 mer TTAAAAA 32% conserved"
repeat_region	3713..3731 /note="2.1 copies 9 mer TTGTGACTT 38% conserved"
repeat_region	3945..3955 /note="2.8 copies 4 mer CATG 22% conserved"
repeat_region	3953..3964 /note="2.0 copies 6 mer CATACA 24% conserved"
repeat_region	3958..4048 /note="22.8 copies 4 mer ACAA 137% conserved"
repeat_region	4048..4093 /note="Dr000319 repeat: matches 184..232 of consensus"
repeat_region	complement(4075)..4252 /note="Dr000027 repeat: matches 5..202 of consensus"
repeat_region	4455..4346 /note="Dr000340 repeat: matches 100..191 of consensus"
repeat_region	4351..4384 /note="2.1 copies 16 mer ATTTATCAAGTGTA 68% conserved"
repeat_region	4738..4751 /note="3.5 copies 4 mer TTGT 28% conserved"
repeat_region	4985..5148 /note="4.1 copies 40 mer ACACGCTGATTCATATACTGAATCGCATGCATGAACAC 138% conserved"
repeat_region	5133..5144 /note="2.0 copies 6 mer CATATA 24% conserved"
repeat_region	5186..5210 /note="2.8 copies 9 mer CATTATTAT 34% conserved"
repeat_region	5195..5217 /note="7.7 copies 3 mer ATT 28% conserved"
repeat_region	5601..5613 /note="3.2 copies 4 mer AAAT 26% conserved"
repeat_region	5709..5722 /note="2.0 copies 7 mer TACAAA 28% conserved"
repeat_region	5743..5765 /note="1.9 copies 12 mer TACTTAAATAC 46% conserved"
repeat_region	5758..5773 /note="2.0 copies 8 mer TTTAAATA 32% conserved"
repeat_region	5862..5878 /note="2.4 copies 7 mer ATAAAT 25% conserved"
repeat_region	5912..5930 /note="2.1 copies 9 mer TTGTGACTT 38% conserved"
repeat_region	6140..6187 /note="3.4 copies 14 mer TACACATACATACA 60% conserved"
repeat_region	6144..6154 /note="2.8 copies 4 mer CATG 22% conserved"
repeat_region	6152..6163 /note="2.0 copies 6 mer CATACA 24% conserved"
repeat_region	6157..6187 /note="7.8 copies 4 mer ACAAT 62% conserved"
repeat_region	complement(6187)..6391 /note="Dr000027 repeat: matches 5..256 of consensus"
repeat_region	6395..6520 /note="Dr000340 repeat: matches 81..207 of consensus"
repeat_region	complement(6449)..6538 /note="Dr000340 repeat: matches 117..207 of consensus"
misc_feature	6449..6522 /note=" tandem repeat. Inconsistency in the number of copies of the repeat element between subclones."
repeat_region	6542..6575 /note="2.1 copies 16 mer ATTTATCAAGTGTA 68% conserved"
repeat_region	6929..6939

repeat_region	/note="2.8 copies 4 mer TTTG 22% conserved"
repeat_region	/note="2.0 copies 6 mer CATATA 24% conserved"
repeat_region	/note="2.8 copies 9 mer CATATTATT 34% conserved"
repeat_region	/note="7.7 copies 3 mer ATT 28% conserved"
repeat_region	/note="2.0 copies 10 mer AGTTTACAT 40% conserved"
repeat_region	/note="4.7 copies 3 mer AAT 28% conserved"
repeat_region	/note="2.0 copies 7 mer TACAAA 28% conserved"
repeat_region	/note="1.9 copies 12 mer TACTTAAATAC 46% conserved"
repeat_region	/note="2.0 copies 8 mer TTTAAATA 32% conserved"

Query Match	Similarity	5.7%	Score 38.2	DB 5	Length 80828
Best Local	Similarity	5.7%	Pred. No. 11		
Matches	73	Conservative	0	Mismatches	58
				Indels	0
				Gaps	0

Query	330	CAGGCGATGCAACCCCTTTGGTTGGAGAGATCTTCATCCATCGATTCATTA	ACT 389
Db	20476	CTGACCATATATTTCTTTTCTGTGAGACATGATATCTCCAGTTGTAACTTCTTGT	20417
Qy	390	CTTGATCCATCTTCTGATGTTTTCAGTGTAAATTAAGAGATGCATGATATATA	449
Db	20416	ATTGGAATAGCCTGTAAAGTTTATGTTGATTACATATCAATAAATATGTACAT	20357
Qy	450	AAATAAGTAA 460	
Db	20356	TAAATTTTAA 20346	

LOCUS	AC104598	126118 bp	DNA	linear	PRI 09-MAR-2002
DEFINITION	Homo sapiens BAC clone Rpi1-182K3 from 2, complete sequence.				
ACCESSION	AC104598				
VERSION	AC104598.4	GI:18693549			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 126118)				
AUTHORS	Sulston, J.E. and Waterston, R.				
TITLE	Toward a complete human genome sequence				
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)				
REFERENCE	99063792				
PUBMED	9847074				
REFERENCE	2 (bases 1 to 126118)				
AUTHORS	Belicki, J., Meyer, R., Haglund, K. and Doeber, A.				
TITLE	The sequence of Homo sapiens BAC clone Rpi1-182K3				
JOURNAL	Unpublished (2001)				
REFERENCE	3 (bases 1 to 126118)				
AUTHORS	Waterston, R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-DEC-2001) Genome Sequencing Center, Washington				
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
AUTHORS	MO 63108, USA				
TITLE	4 (bases 1 to 126118)				
JOURNAL	Waterston, R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (16-FEB-2002) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
JOURNAL	MO 63108, USA				
REFERENCE	5 (bases 1 to 126118)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-MAR-2002) Department of Genetics, Washington				

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 16, 2002 this sequence version replaced qi:18250125.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/qsc>

Web site: <http://journals.wustl.edu>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

===== Summary Statistics =====

----- summary statistics  
center project name: H NH0182K03

Center proj

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**

SOURCE: Immunization BAC liquid was made from the blood of one male (donor, M. Catangen, J and de Jong P.J. (1998) An improved technique for construction of bacterial artificial chromosome libraries Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBAC3.6

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-252K7, 2000 bp overlap.  
The clone sequenced to the right is RP11-182K3.  
Actual start of this clone is at base position 106699 of  
RP11-252K7. Actual end is at base position 126118 of RP11-182K3.

## FEATURES

Source

1. .126118	
organism="Homo sapiens"	
db_xref="taxon:9606"	
chromosome="2"	
map="2"	
clone="RP11-182K3"	
clone.lib="RPC1-11"	
892. .1918	
rpt_family="(A)n"	
0003. .2229	
note="similar to Homo sapiens EST BF890821 (NID:gl2282280)"	
4309. .5520	
rpt_family="L2"	
6505. .5990	
rpt_family="MaLR"	
6191. .6700	
note="match to EST AL037617 (NID:g5928218)"	
6596. .7193	
rpt_family="L2"	
7525. .7604	
note="match to EST AL037954 (NID:g5407270)"	
7525. .7604	
note="match to EST BE811024 (NID:g10243258)"	
7525. .7604	
note="match to EST BE811026 (NID:g10243260)"	
7525. .7604	
note="match to EST BE926756 (NID:g10452832)"	
10223. .10361	note="match to EST AL517947 (NID:g12781440)"
10319. .10361	note="similar to Bos taurus EST BE810028 (NID:g10241140)"
11371. .11466	rpt_family="MER2_type"
11467. .11735	rpt_family="MER2_type"
11736. .12026	rpt_family="Alu"
12027. .12481	rpt_family="MER2_type"
12614. .12774	note="match to EST BG983142 (NID:g14385864)"
12733. .13000	rpt_family="Alu"
13003. .13026	rpt_family="(TG)n"
13162. .13187	rpt_family="AT-rich"
13615. .13994	note="match to EST AA328955 (NID:g1981442)"
13679. .14248	note="match to EST AI950302 (NID:g5742612) wp08b02.x1"
13841. .14227	note="match to EST R93169 (NID:g967335) yt69c02.e1"
13884. .14235	note="match to EST AA682675 (NID:g2669956) zj20h11.e1"

```

repeat_region 14236..14256
                /rpt_family="(A)n"
misc_feature 1438..1479
                /note="match to EST R82015 (NID:g858618) y106f08.r1"
repeat_region 15005..15305
                /rpt_family="Alu"
misc_feature 17481..17958
                /note="match to EST BF771506 (NID:g12119406)"
repeat_region 17786..18083
                /rpt_family="Alu"
repeat_region 18084..18129
                /rpt_family="(TCTA)n"
repeat_region 18691..18712
                /rpt_family="AT_rich"
misc_feature 19094..19636
                /note="match to EST AL037616 (NID:g5928217)"
repeat_region 19686..19709
                /rpt_family="AT_rich"
repeat_region 19804..19825

```

Query Match 7.7%; Score 38.2; DB 9; Length 126118;

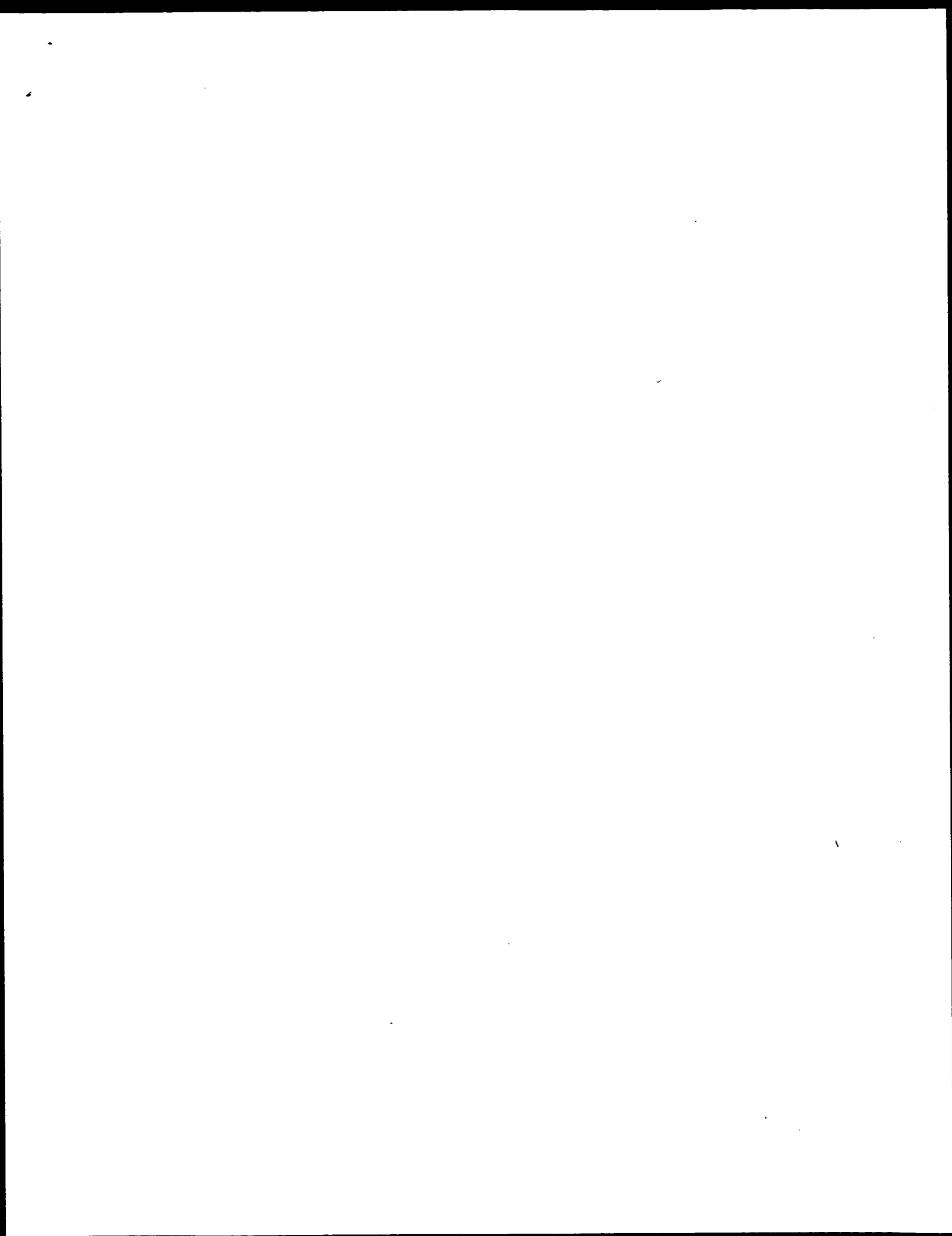
Best Local Similarity 53.7%; Pred. No. 10; Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

QY 331 AGGCGATGCACCCCTTTGGTTGAGAGATATCTTCATCCATGCTAGATTTCATATCTC 390
    |||||
Db 59680 AGTCCTACCCCGTTGCATCATGATTAAGTACATCATGTGCTGTAAGTT 59739
    |||||
QY 391 TTGATCCATCTTCTATGTTTTCAGTGTATTAATTAGAGATGATGATATATATA 450
    |||||
Db 59740 TTGGTGCTTCTGCTGACTTTATATCTATTTATGAGGACGTGAATAGCTGATTTAT 59799
    |||||
QY 451 AATAGTAAAGCTACGATACCAT 477
    |||||
Db 59800 ATGAATTCAAATGTGAGCAAAAAAAT 59826
    |||||

```

Search completed: January 12, 2003, 08:14:38  
Job time : 3376 secs















PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232568.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239933.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

XX Disclosure; SEQ ID NO 39112; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

XX SQ Sequence 8209 BP; 2721 A; 1260 C; 1380 G; 2848 T; 0 other;

Query Match

Best Local Similarity 7.2%; Score 35.4; DB 22; Length 8209;

Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 345 TTTTGGTTGAAGAGTATCTTCATCAATGCTAGATTTTCATTAACCTTTGGATCCATCTTC 404

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2563 TATTGTAGATGCACTCACTAATTTTTCGATTTCGAACATTTTTC 2622  
 QY 405 TATGTTTTCAAGTGTATTAATAGAGATGATATATAATTAAGTAAAGCT 464  
 Db 2623 TCACCTGTTTCATTCGAGATATATGAATATCATTCATTTTAAATATAAGCAATTCCT 2682  
 QY 465 ACGGT 469  
 Db 2683 ACTGT 2687

RESULT 8  
 AAK84301  
 ID AAK84301 standard; DNA, 8210 BP.  
 XX  
 AC AAK84301;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39113.  
 XX  
 KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KM Cytotoxic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01354.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184654.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225277.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226679.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0232403.  
 PR 14-SEP-2000; 2000US-0232404.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241122.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 01-NOV-2000; 2000US-0241826.  
 PR 08-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
XX Disclosure; SEQ ID NO 39113; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/haematopoietic-related diseases, especially  
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/haematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
XX represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 8210 BP; 2722 A; 1260 C; 1380 G; 2848 T; 0 other;  
XX  
XX Query Match 7.2%; Score 35.4; DB 22; Length 8210;  
XX Best Local Similarity 55.2%; Pred. No. 2.7; Indels 0; Gaps 0;  
XX Matches 69; Conservative 0; Mismatches 56;  
XX  
XX 345 TTTTGGTTGGAAGAGTATCTTCATCCAAATGCTAGATTTCATCACTCTTGGATCCATCTTC 404  
XX 2563 TATTTGTAGGATGGCACTCATCTATTATAATTTTTTTTGGCATTTGCCAACATTTTTC 2622  
XX  
XX 405 TATGTTTTTCAAGTGTATTAATAGAGAGATGCATGATATATAATAAGTAAAGCT 464  
XX 2623 TCACCTTGTTCATTTTCGAGGATATATGAATGCATTGATTTTAAATATAAAGCGAATCT 2682  
XX  
XX 465 ACGGT 469  
XX 2683 ACTGT 2687  
XX  
XX RESULT 9

ABL34024  
ID ABL34024 standard; DNA; 6375 BP.  
XX  
AC ABL34024;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 1997.  
XX  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
XX for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation -  
XX  
XX Claim 1; SEQ ID NO 1997; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 6375 BP; 1795 A; 207 C; 1431 G; 2942 T; 0 other;  
XX  
XX Query Match 7.1%; Score 35.2; DB 24; Length 6375;  
XX Best Local Similarity 55.8%; Pred. No. 2.7; Indels 0; Gaps 0;  
XX Matches 67; Conservative 0; Mismatches 53;  
XX  
XX 346 TTTGGTTGGAAGAGTATCTTCATCCAAATGCTAGATTTCATCACTCTTGGATCCATCTTC 405  
XX 2241 TTTTGTGTAAGATTATTTTATTTATGCGGGTTTATTAGTTTTTTTATTTATGTTGA 2300  
XX  
XX 406 ATGTTTTTCAAGTGTATTAATAGAGAGATGCATGATATATAATAAGTAAAGCTA 465  
XX 2301 TTATAAAAAATGTTTTTTTATTAGATAAAATTTATTATAAAAAATAGATTATTATAAGATA 2360  
XX  
XX RESULT 10  
XX AAS46745  
XX ID AAS46745 standard; DNA; 38342 BP.  
XX  
XX AAS46745;  
XX  
XX 18-DEC-2001 (first entry)  
XX

DE Tumour suppressor gene derived chemically modified sequence #469.  
XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic;  
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
XX cytosine methylation; ds.  
OS Homo sapiens.  
XX  
XX MO200168912-A2.  
XX  
XX  
XX 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP02955.  
XX  
XX 15-MAR-2000; 2000DE-1013847.  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2001-602752/68.  
XX  
XX  
XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer -  
XX  
XX  
XX Claim 1; SEQ ID No 469; 27bp; English.  
XX  
XX The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (58) and sequences complementary to (58). The nucleic acid may be a  
CC polypeptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 38342 BP; 1153 A; 472 C; 7565 G; 19152 T; 0 other;  
Query Match 7.1%; Score 35.2; DB 22; Length 38342;  
Best Local Similarity 55.8%; Pred. No. 6.8;  
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
OY 346 TTTGGTTGGAAGATCTTCATCCAAATGCTAGATTTCATACCTTGGATCCATCTTCT 405  
DB 2241 TTTTGTGTAAGATTTATTTTATTTATGCGGGTTTATTTAGTTTATTTATGCTGA 2300  
OY 406 ATGTTTTCAGGTATTAATTAGAGATGATGATATTAATTAATAAGTAAAGCTA 465  
DB 2301 TTATAAAAAATGTTTATTTATTTAGATAAATTTATTAATAAAAAATTAAGTTTAAAGATA 2360  
RESULT 11

ABK31506  
ID ABK31506 standard; DNA; 38342 BP.  
XX  
XX  
XX AC ABK31506;  
XX  
XX  
XX 23-APR-2002 (first entry)  
XX  
XX  
XX Signal transduction associated gene modified DNA #175.  
DE  
XX  
XX  
XX Human; signal transduction associated gene; cytosine methylation state;  
KW CpG island; signal transduction associated disease; solid tumour; cancer;  
XX antitumour; cytosolic; mutant; ds.  
XX  
XX  
XX Homo sapiens.  
OS  
OS Synthetic.  
XX  
XX  
XX MO200200926-A2.  
XX  
XX  
XX 03-JAN-2002.  
XX  
XX  
XX 29-JUN-2001; 2001WO-EP07472.  
XX  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-147896/19.  
XX  
XX  
XX Oligonucleotide for diagnosis and therapy of diseases associated with  
PT signal transduction e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with signal transduction -  
XX  
XX  
XX Claim 1; SEQ ID No 349; 24bp; English.  
XX  
XX  
XX The present invention relates to chemically modified DNA sequences of  
CC signal transduction associated genes. The DNA sequences are chemically  
CC modified using a solution of bisulphite, hydrogen sulphite or  
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
CC for detecting the cytosine methylation state (CpG islands) of these  
CC genes, and a method for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with signal transduction.  
CC The genomic DNA can be obtained from cells or cellular components which  
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from  
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
CC histologic object slides, and all their possible combinations. The  
CC sequences of the invention are useful for the diagnosis and therapy of  
CC diseases associated with signal transduction e.g. solid tumours and  
CC cancer. ABK3158-ABK31545 represent chemically pretreated genomic DNA  
CC sequences of different genes associated with signal transduction, or  
CC their complementary sequences.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX  
SQ Sequence 38342 BP; 1153 A; 472 C; 7565 G; 19152 T; 0 other;  
Query Match 7.1%; Score 35.2; DB 24; Length 38342;  
Best Local Similarity 55.8%; Pred. No. 6.8;  
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
OY 346 TTTGGTTGGAAGATCTTCATCCAAATGCTAGATTTCATACCTTGGATCCATCTTCT 405  
DB 2241 TTTTGTGTAAGATTTATTTTATTTATGCGGGTTTATTTAGTTTATTTATGCTGA 2300  
OY 406 ATGTTTTCAGGTATTAATTAGAGATGATGATATTAATTAATAAGTAAAGCTA 465  
DB 2301 TTATAAAAAATGTTTATTTATTTAGATAAATTTATTAATAAAAAATTAAGTTTAAAGATA 2360

Db 95 AAGTTCTTCTGTACTC 78

RESULT 13  
 AAX82876/c  
 ID AAX82876 standard; DNA; 179 BP.  
 XX  
 XX AAX82876;  
 XX  
 DT 30-JUN-2000 (first entry).  
 XX  
 DE Human dysferlin DNA #24.  
 XX  
 DE Dysferlin; anti-dystrophic; gene therapy; muscular dystrophy; human;  
 KW skeletal muscle cell; hereditary; Miyoshi myopathy; diagnosis;  
 KW limb girdle muscular dystrophy-2B; brain-specific; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 WO WO200011157-A1.  
 PN  
 PD 02-MAR-2000.  
 XX  
 XX 25-AUG-1999; 99WO-US19395.  
 XX  
 XX 25-AUG-1998; 98US-0097927.  
 XX  
 XX (GEO ) GEN HOSPITAL CORP.  
 XX  
 PI Brown RH, Liu J, Aoki M, Ho MF, Mateuda-Asada C;  
 XX  
 DR WPI; 2000-237646/20.  
 XX  
 PT Novel dysferlin genes and related proteins useful for diagnosis, risk  
 PT identification and treatment of hereditary muscular dystrophies and  
 PT other dysferlin related disorders -  
 XX  
 XX Claim 11; Page 111; 146pp; English.  
 XX  
 XX This invention describes a novel human dysferlin nucleic acid (I) and  
 CC its encoding protein (II), which has anti-dystrophic activity and can be  
 CC used for gene therapy. Introduction of (I), a vector comprising (I) or  
 CC dysferlin into a cell of a mammal can be used to decrease the symptoms of  
 CC muscular dystrophy. The dysferlin gene is normally expressed in skeletal  
 CC muscle cells and is selectively mutated in several families with the  
 CC hereditary muscular dystrophies, e.g. Miyoshi myopathy and limb girdle  
 CC muscular dystrophy-2B. The primers and oligonucleotides derived from (I)  
 CC can be used in diagnosis of or risk identification for dysferlin-related  
 CC disorders in patients, fetus, or pre-embryos. Expression of  
 CC brain-specific dysferlin may be important as a marker for normal neural  
 CC development. Dysferlin DNA or subgenomic coding sequences can be used for  
 CC therapy of the hereditary muscular dystrophies. This sequence represents  
 CC a fragment of the human dysferlin gene described in the method of the  
 XX invention.  
 XX  
 SQ Sequence 179 BP; 42 A; 34 C; 66 G; 37 T; 0 other;

Query Match 7.1%; Score 34.8; DB 21; Length 179;  
 Best Local Similarity 65.4%; Pred. No. 0.6;  
 Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0

QY 22 CATACACTCTTCTCCTCCCAACCAATTAGCACTTATCAGCTTAACCTCAGGCATGGCTTCCACC 81  
 DB 155 CCTGCTCCCTTGCTCTTACCATTTCCTCCGCAAGCTGACCTCCACAAAGGGGTCCACC 96  
 QY 82 AAGTTGTTCTTCTCAGTC 99  
 DB 95 AAGTTCTTCTGTACTC 78

RESULT 14  
 NAv87102









QY 420 TATAATTAGAGATGATGATATATAATAAGTAAAGTACGGTATCACCATGT 479  
Db 4911 TCATATTAGAAATAGCTTTGTAATCATAGAAAGAGGTGTATGCAATGAATTAAGTCT 4852  
QY 480 GATGATTTTACC 492  
Db 4851 TGTGAAGCTACC 4839

## RESULT 2

US-09-134-001C-1936/c  
; Sequence 1936, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1936  
; LENGTH: 1422  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1936

Query Match 6.5%; Score 32.2; DB 4; Length 1422;  
Best Local Similarity 57.4%; Pred. No. 1.1;  
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 371 AATGCTAGATTTTATCAATCTTGGATCCATCTTCTATGTTTTCAGTGATATAATTAGAG 430  
Db 1232 AATCTCTTTGTTTAACTTCAATATATTTGGAGATTAAATCATATTATCTTT 1173  
QY 431 AGATGATGATATATAATAAGTAAAGTACGGTAT 471  
Db 1172 AGATCAAAATATAATAAACTAGTAAAGTCTGTCT 1132

## RESULT 3

US-08-845-539-5  
; Sequence 5, Application US/08845539  
; Patent No. 5929303  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Rose, Jocelyn K.C.  
; TITLE OF INVENTION: Fruit-Specific and Ripening-Regulation  
; TITLE OF INVENTION: Expansin Genes to Control Fruit Texture and Softening  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/845,539  
; FILING DATE: 25-APR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 023070-078200US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 537 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..537  
; OTHER INFORMATION: /product= "melon CmEx1"  
US-08-845-539-5

Query Match 6.5%; Score 31.8; DB 2; Length 537;  
Best Local Similarity 52.7%; Pred. No. 0.88;  
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 238 GGCTATGACTTACGCTACACTGGACAACTGCTCTCTACAAACGAGCTGGATGCAGT 297  
Db 82 GGCTATGGCGTCAACACAGCTGCTCTTAGTACTGCTTTCTTCAACAATGGCCTCAGCTGT 141  
QY 298 GGTGTTGCACACACAGGTTTGGTCCAGTCCAGGCGCATGCAACCCCTTTTGGTTGGAAG 357  
Db 142 GGTGCTTGGCTTTGAGATCAAGTGTCTAATGACCCCTCGATGGTCCATCCTGTTAGCCCT 201  
QY 358 AGTATCTTCAT 368  
Db 202 TGTATCTTCAT 212

## RESULT 4

US-09-362-642-5  
; Sequence 5, Application US/09362642  
; Patent No. 6350935  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Rose, Jocelyn K.C.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Fruit-Specific and Ripening Regulation Expansin Genes  
; TITLE OF INVENTION: to Control Fruit Texture and Softening  
; FILE REFERENCE: 023070-078210US  
; CURRENT APPLICATION NUMBER: US/09/362,642  
; CURRENT FILING DATE: 1999-07-27  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 537  
; TYPE: DNA  
; ORGANISM: Cucumis melo  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(537)  
; OTHER INFORMATION: melon expansin (CmEx1) partial cDNA clone  
US-09-362-642-5

Query Match 6.5%; Score 31.8; DB 4; Length 537;  
Best Local Similarity 52.7%; Pred. No. 0.88;  
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 238 GGCTATGACTTACGCTACACTGGACAACTGCTCTCTACAAACGAGCTGGATGCAGT 297  
Db 82 GGCTATGGCGTCAACACAGCTGCTCTTAGTACTGCTTTCTTCAACAATGGCCTCAGCTGT 141  
QY 298 GGTGTTGCACACACAGGTTTGGTCCAGTCCAGGCGCATGCAACCCCTTTTGGTTGGAAG 357  
Db 142 GGTGCTTGGCTTTGAGATCAAGTGTCTAATGACCCCTCGATGGTCCATCCTGTTAGCCCT 201  
QY 358 AGTATCTTCAT 368  
Db 202 TGTATCTTCAT 212



GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchinashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-20

Query Match 6.4%; Score 31.4; DB 2; Length 246240;  
Best Local Similarity 50.3%; Pred. No. 28;  
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
QY 131 GTGAGATGGTGAATGGGAGTGCATTTACAGTATGGAGTGGTCCAGGTTGTAAACAACCGTG 190  
DB 145036 GGGAGAGGAGGACAGGTGGTTATTTTATGCTCATAGGCTATATATTACACAATAGAGT 144977  
QY 191 CTGAGCGATATAGCAAGTGTGGATGCTCAGCTATACATCAGAGGAGGCTATGACTTCA 250  
DB 144976 CATACATATTTAGCAGCTTTGGGGGACAGCTATATATTATGAGGGTGCCAAAGTGCA 144917  
QY 251 GCTACACTGGACAAACTGCTGCTCTACAACC 283  
DB 144916 TTCACAATGGATAAACACGCTGTAATATACCTCC 144884

RESULT 8  
US-08-724-394A-21/c  
Sequence 21, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston  
APPLICANT: Tsuchinashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-21

Query Match 6.4%; Score 31.4; DB 2; Length 246240;  
Best Local Similarity 50.3%; Pred. No. 28;  
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
QY 131 GTGAGATGGTGAATGGGAGTGCATTTACAGTATGGAGTGGTCCAGGTTGTAAACAACCGTG 190  
DB 145036 GGGAGAGGAGGACAGGTGGTTATTTTATGCTCATAGGCTATATATTACACAATAGAGT 144977  
QY 191 CTGAGCGATATAGCAAGTGTGGATGCTCAGCTATACATCAGAGGAGGCTATGACTTCA 250  
DB 144976 CATACATATTTAGCAGCTTTGGGGGACAGCTATATATTATGAGGGTGCCAAAGTGCA 144917  
QY 251 GCTACACTGGACAAACTGCTGCTCTACAACC 283  
DB 144916 TTCACAATGGATAAACACGCTGTAATATACCTCC 144884

RESULT 9  
US-08-724-394A-22/c  
Sequence 22, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchinashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-21

NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Flets, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "H1A-H-CONTIG"  
US-08-724-394A-22

Query Match  
Best Local Similarity 50.3%; Score 31.4; DB 2; Length 246240;  
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 131 GTGAGTGTGGAATGGAGTGCATTTTACATGATGAGTGGTCCAGGTTGTAACACCGTG 190  
DB 145036 GGGAGAGAGAGCAGCTGTTATTTTATGCTCATAGGCTATATTATTAACAATGAGT 144977  
QY 191 CTAGAGCATATAGCAAGTGTGATGCTCAGCTATACATCAGAGGAGGCTATGACTTCA 250  
DB 144976 CATACATATTAGCAGCTTTGGGGGAGCAGCTATATATTATGAGGGGTCACAACTGCA 144917  
QY 251 GCTACACTGACAACTGCTGCTCTCTACACC 283  
DB 144916 TTCACATGATTAACACGCTATATATACCTCC 144884

RESULT 10  
US-07-741-453A-57/c  
Sequence 57, Application US/07741453A  
Patent No. 6228597  
GENERAL INFORMATION:  
APPLICANT: PARMENTIER, MARC  
APPLICANT: LIBERT, FREDERIC  
APPLICANT: DUMONT, JACQUES  
APPLICANT: VASSART, GILBERT  
TITLE OF INVENTION: POLYPEPTIDES HAVING THYROTROPIN-RECEPTOR  
TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SUCH RECEPTORS  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DABRY & CUSHMAN  
STREET: 1615 L STREET, N.W.  
CITY: WASHINGTON, D.C.  
COUNTRY: U.S.A.  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/741,453A  
FILING DATE: 19911015  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16773  
REFERENCE/DOCKET NUMBER: 91913/1107/US/57  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4417 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-07-741-453A-57

Query Match  
Best Local Similarity 45.8%; Score 31.2; DB 4; Length 4417;  
Matches 108; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 85 TTGTTCTTCTCAGTCATTACTGTGTGATGCTCATAGCAATGGCAAGTGAATGATG 144  
DB 499 TTGTCACATCAGAGAAATCTCAAGTCCAGTGTGAATAATGCCAAGAACTTCAGAGCT 440  
QY 145 GGGAGTGAATTACGATAGAGTGTCCAGCTGTAAACCGTGTGACGATATAGC 204  
DB 439 GGGAGCTCTTTTGGGCGCTCAGGCTCTATGATGTTAAGCTTCTGTAATTCGATCTCT 380  
QY 205 AAGTGTGATGCTCAGTATACATCAGAGGAGGCTATGACTTCAGCTACAGTACGACAA 264  
DB 379 AGCTAGCATTTTCTTAATTTGAGAGAAATGTGATTCAGCCGCTGCAAGATTGCA 320  
QY 265 ACTGCTGCTCTTACACACAGGCTGATGATGAGTGTGGACACACACAGGTTTG 320  
DB 319 TCTATTGACAAAGTATAGCTCGAAATATTGGGAGATTTGAAATGACGACTGG 264

RESULT 11  
US-08-121-446-3/c  
Sequence 3, Application US/08121446  
Patent No. 6313276  
GENERAL INFORMATION:  
APPLICANT: IMURA, HIROO  
APPLICANT: NAKAO, KAZUWA  
APPLICANT: NAKANISHI, SHIGETADA  
TITLE OF INVENTION: A HUMAN ENDOTHELIN RECEPTOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/121,446  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/911,684  
; FILING DATE: 10-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CIOTTI, THOMAS E.  
; REGISTRATION NUMBER: 21,013  
; REFERENCE/DOCKET NUMBER: 29900-20324.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4301 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 238..1566  
US-08-121-446-3

Query Match 6.3%; Score 31; DB 4; Length 4301;  
Best Local Similarity 51.0%; Pred. No. 4.7; Mismatches 70; Indels 0; Gaps 0;  
Matches 73; Conservative 0;  
QY 344 CTTTGGTTGAAGAGTATCTTCATCCCAATGCTAGATTTCATACTCTTGGATCCATCTT 403  
DB 3673 CATTTTITAAAAGCCCACTGAATGCAATTTTATTATAAATAATGCAAGTATGCTTTT 3614  
QY 404 CTATGTTTTCAAGTGTATAATTAGAGAGATGCATGGATATATAATAAATGAAGTAAAGC 463  
DB 3613 GTGTGTAGTTTCAAAATTTGGCAAGATAATTAGAGAACATGATATCTGACATGGAGAAAGGA 3554  
QY 464 TACGGTATCACCATGTGATTT 486  
DB 3553 AAGGGTATCAGGATGTAATAATT 3531

RESULT 12  
US-09-741-150-3/c  
; Sequence 3, Application US/09741150  
; Patent No. 6436689  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO00968  
; CURRENT APPLICATION NUMBER: US/09/741,150  
; CURRENT FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 112132  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: (1)...(112132)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-741-150-3

Query Match 6.3%; Score 31; DB 4; Length 112132;  
Best Local Similarity 56.3%; Pred. No. 26; Mismatches 45; Indels 0; Gaps 0;  
Matches 58; Conservative 0;  
QY 371 AATGCTAGATTTCAATACTCTTGGATCCATCTTCTATGTTTTTCAAGTGTATAATTAGAG 430  
DB 106795 AATGTGAAAAATGTGCAGTTTCTTTTATTGATCAAAAATTTAAATATGATGATAGTT 106736  
QY 431 AGATCGATGATATATAATAAATAAGTAAAGCTACGGTATCA 473

DB 106735 ACAGGCATAAAATATAATATATATACTTAAACCAAGGTTTTCA 106693

RESULT 13  
US-09-600-776-9  
; Sequence 9, Application US/09600776  
; Patent No. 6326168  
; GENERAL INFORMATION:  
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.  
; TITLE OF INVENTION: A novel potassium channel protein  
; FILE REFERENCE: Y9903-PCT  
; CURRENT APPLICATION NUMBER: US/09/600,776  
; CURRENT FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: JP P1998-011434  
; PRIOR FILING DATE: 1998-01-23  
; PRIOR APPLICATION NUMBER: JP P1998-346198  
; PRIOR FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 3715  
; TYPE: DNA  
; ORGANISM: Rattus sp.  
US-09-600-776-9

Query Match 6.2%; Score 30.8; DB 4; Length 3715;  
Best Local Similarity 61.0%; Pred. No. 5.1; Mismatches 32; Indels 0; Gaps 0;  
Matches 50; Conservative 0;  
QY 233 AGGAGGCTATGACTTCAGCTACACTGGACAAACCTGCTCTCTACAAACAGGCTGGAT 292  
DB 2772 AGTGAGGCAAGAAACACACACACTGGCAAGCTACGGCAGCGGTGACGGAGCTGTCT 2831  
QY 293 GCAGTGGTGTTCACACACCAG 314  
DB 2832 GAACAGGTGCTGCAGATGCGAG 2853

RESULT 14  
US-08-973-462-2/c  
; Sequence 2, Application US/08973462B  
; Patent No. 6191270  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIERRE  
; APPLICANT: DAUBERSIES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 0660-0125-0 PCT  
; CURRENT APPLICATION NUMBER: US/08/973,462B  
; CURRENT FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
; EARLIER FILING DATE: 1996-06-12  
; EARLIER APPLICATION NUMBER: FR 95/07007  
; EARLIER FILING DATE: 1995-06-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 5361  
; TYPE: DNA  
; ORGANISM: P. falciparum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(5361)  
US-08-973-462-2

Query Match 6.2%; Score 30.6; DB 4; Length 5361;  
Best Local Similarity 52.0%; Pred. No. 7.2; Mismatches 60; Indels 0; Gaps 0;  
Matches 66; Conservative 1;  
QY 364 TTCATCAATGCTAGATTTCATCAACTCTTGGATCCATCTTCTATGTTTTTCAAGTGATA 423  
DB 5244 TACTTCTTATACATTATATATTTTTTTCTTACTAAATTTTTTCTTCAATTTGTTGTAATA 5185  
QY 424 ATTAGAGATGATGATGATATATAATAAATAAGTAAAGCTACGGTATCACCATGTGATG 483

Db 5184 ATGTTTAAATCTTAAAGATATCAAAATCAATGAATAAACGGTCTCTCTTATTTT 5125  
QY 484 ATTTTA 490  
Db 5124 ATTTATA 5118

## RESULT 15

US-08-973-462-1/c  
; Sequence 1, Application US/08973462B  
; Patent No. 6191270  
; GENERAL INFORMATION:  
; APPLICANT: DRUIHE, PIERRE  
; APPLICANT: DAUBERES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 0660-0125-0 PCT  
; CURRENT APPLICATION NUMBER: US/08/973,462B  
; CURRENT FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
; EARLIER FILING DATE: 1996-06-12  
; EARLIER APPLICATION NUMBER: FR 95/07007  
; EARLIER FILING DATE: 1995-06-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 6152  
; TYPE: DNA  
; ORGANISM: P. falciparum  
US-08-973-462-1

## Query Match

Best Local Similarity 52.0%; Score 30.6; DB 4; Length 6152;  
Matches 66; Conservative 1; Mismatches 60; Indels 0; Gaps 0;

QY 364 TTCATCCAAATGCTAGATTTCATTAACCTCTGGATCCATCTTCTATGTTTCAAGTGATA 423  
Db 5488 TACTTCTATTAATTAATTTTCTTACTTAATTTTCTTCAATTTGTTGAATA 5429  
QY 424 ATTAGAGAGATGATGATATATATAATTAAGCTACGATACCAATGTGATG 483  
Db 5428 ATGTTTAAATCTTAAAGATATCAAAATCAATGAATAAACGGTCTCTCTTATTTT 5369  
QY 484 ATTTTA 490  
Db 5168 ATTTATA 5362

Search completed: January 12, 2003, 08:42:01  
Job time : 365 secs







Db 61 ACCTCAGCCATGGCTTCACCAAGTTGTTCTTCAGTCAATTAAGTCAATGCTCATA 120  
Qy 121 GCAATGGCAAGTGAAGTGGTGAATGGAGTGCATTTACAGTATGAGTGGTCCAGGTTGT 180  
Db 121 GCAATGGCAAGTGAAGTGGTGAATGGAGTGCATTTACAGTATGAGTGGTCCAGGTTGT 180  
Qy 181 AACAAACCGTGTGAGCGGATATAGCAAGTGTGGATGCTCAGCTATACATCAGAGGGAGGC 240  
Db 181 AACAAACCGTGTGAGCGGATATAGCAAGTGTGGATGCTCAGCTATACATCAGAGGGAGGC 240  
Qy 241 TATGACTTCAGCTACACTGGAACAACCTGCTCTCTACAAACAGGCTGGATGCAAGTGT 300  
Db 241 TATGACTTCAGCTACACTGGAACAACCTGCTCTCTACAAACAGGCTGGATGCAAGTGT 300  
Qy 301 GTTGCACACACAGGTTTGGGTCAGTGCAGGCGCATGCCAACCCCTTTGGTTGGAAGAGT 360  
Db 301 GTTGCACACACAGGTTTGGGTCAGTGCAGGCGCATGCCAACCCCTTTGGTTGGAAGAGT 360  
Qy 361 ATCTTCATCCAAATGCTAGATTTCACTCTCTGGATCCATCTTCTATGTTTTCAGTGT 420  
Db 361 ATCTTCATCCAAATGCTAGATTTCACTCTCTGGATCCATCTTCTATGTTTTCAGTGT 420  
Qy 421 ATAATTAGAGAGTGCATGGATATATAATAAATAAGTAAAGCTACCGTATCACCATGTG 480  
Db 421 ATAATTAGAGAGTGCATGGATATATAATAAATAAGTAAAGCTACCGTATCACCATGTG 480  
Qy 481 ATGATTTTACCC 493  
Db 481 ATGATTTTACCC 493

RESULT 2  
US-09-864-761-6488  
; Sequence 6488, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 6488  
; LENGTH: 463  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC003025.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5  
US-09-864-761-6488

Query Match 6.6%; Score 32.4; DB 10; Length 463;  
Best Local Similarity 60.0%; Pred. No. 2.4;  
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
Qy 264 AACTGCTGCTCTCTACAACAGGCTGGATGAGTGTGTCACACACAGGTTGGTGC 323  
Db 152 AGCTGCAGCTGTGTGCAGCCAGGAGGAGTTCTCTGAGCACATGCTAGGTGCTGCGC 211  
Qy 324 CAGTCCAGGCGATCAACCCCTTTTGGTTG 353  
Db 212 CAGTCCAGGCGAGGAGGATATAATAAAGTTG 241

RESULT 3  
US-09-734-674-3/c  
; Sequence 3, Application US/09734674  
; Patent No. US20020081648A1  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ming-Hui et al  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001018  
; CURRENT APPLICATION NUMBER: US/09/734,674  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 202001  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(202001)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-734-674-3

Query Match 6.5%; Score 32; DB 10; Length 202001;  
Best Local Similarity 56.7%; Pred. No. 77;  
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
Qy 10 TTGAGTCTCATACATACCTCTCTCTCCCTCCACCATAGACATTATCAGCTAACCTCAGCC 69  
Db 150243 TTGAGTTGAATACCTGATTTCTGCTCCCTTTCTAGCTATTGATCTTAGCAATAACCCCAACC 150184  
Qy 70 ATGGCTTCCACCAAGTTGTTCTTCTCAGTCATTACTGTGATGAT 113



```
Db 293 TGGGTTTTCTGGTGTCAATA 273
US-09-764-864-604/c
; Sequence 604, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 604
; LENGTH: 638
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-864-604

Query Match 6.4%; Score 31.4; DB 10; Length 638;
Best Local Similarity 61.7%; Pred. No. 5.9;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 228 TCAGAGGAGGCTATGACTTACGTACACTGGACAACTGCTCTCTACACACGAGC 287
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 TCAGAGAGTTGGTACAACTTCTGCTGCACTGAACACTGGTAGGTTCTCTCAGCA 346
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 288 TGGATGACAGTGGTGTTCACA 308
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 TGGGTTTTCTGGTGTCAATA 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-09-754-853A-2/c
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763)..(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-3

Query Match 6.3%; Score 31.2; DB 9; Length 335913;
Best Local Similarity 52.4%; Pred. No. 1.7e+02;
Matches 66; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

Qy 364 TTCATCCAATGCTAGATTTTCATAAATCTTGGATCCATCTTCTATGTTTTCACAGTGATA 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185216 TTTTTCATAAGCTAGAAATTTCTATCTATTAGTTTCTTCAAAAAATTTATTTCTAACTTATG 185157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 424 ATTAGAGAGATGCATGGATATATAATAAAGTAAAGCTACGGTATCACCATGTGATG 483
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185156 AATAAACTAAATTTTAGTTTATATAAAAAAATTTTCAATTTTTTCTTACAAGCATTTATAA 185097
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 484 ATTTTY 489
|||||
Db 185096 ATTTT 185091
|||||

RESULT 10
US-09-754-853A-3/c
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763)..(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-3

Query Match 6.3%; Score 31.2; DB 9; Length 335913;
Best Local Similarity 52.4%; Pred. No. 1.7e+02;
Matches 66; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

Qy 364 TTCATCCAATGCTAGATTTTCATAAATCTTGGATCCATCTTCTATGTTTTCACAGTGATA 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185216 TTTTTCATAAGCTAGAAATTTCTATCTATTAGTTTCTTCAAAAAATTTATTTCTAACTTATG 185157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 424 ATTAGAGAGATGCATGGATATATAATAAAGTAAAGCTACGGTATCACCATGTGATG 483
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185156 AATAAACTAAATTTTAGTTTATATAAAAAAATTTTCAATTTTTTCTTACAAGCATTTATAA 185097
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 484 ATTTTY 489
|||||
Db 185096 ATTTT 185091
|||||

RESULT 11
US-09-931-157-2/c
; Sequence 2, Application US/099311157
; Patent No. US20020082414A1
; GENERAL INFORMATION:
; APPLICANT: Imura, Hiroo
; APPLICANT: Nakao, Kazuwa
; APPLICANT: Nakanishi, Shigetada
; TITLE OF INVENTION: Human Endothelin Receptor
; FILE REFERENCE: 299002032411
; CURRENT APPLICATION NUMBER: US/09/931,157
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 08/121,446
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 07/911,684
; PRIOR FILING DATE: 1992-07-10
```

```

; PRIOR APPLICATION NUMBER: JP 3-172828
; PRIOR FILING DATE: 1991-07-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 4301
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (238) ... (1566)
US-09-931-157-2

Query Match
Best Local Similarity 6.3%; Score 31; DB 10; Length 4301;
Pred. No. 21;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 344 CTTTGGTGAAGATGATCTTCATGCTAGTATTCATCTTGGATGATCTT 403
DB 3673 CATTTTAAAAAGCAGCTGATGATTTTATTAATGCAAGTATGCTTTT 3614
QY 404 CTATGTTTTCAGTATATTAATGAGATGATGATATATTAATGATAAGC 463
DB 3613 GTGTAGTTTCAAAATTTGGCAAGATTAATGAGAAATGATATCTGATGAGAAAGCA 3554
QY 464 TACGCTATCAGCATGTGATGATT 486
DB 3553 AAGGATACAGATGATAAAATT 3531

RESULT 12
US-09-882-434a-2/c
; Sequence 2, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulet, Kenneth C.
; APPLICANT: Green, Jodie Lyn
; TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
; FILE REFERENCE: CULN18.1CP1C1
; CURRENT APPLICATION NUMBER: US/09/882,434A
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/364395
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/117615
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Macadamia integrifolia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70) ... (375)
; OTHER INFORMATION: y=t or c.
US-09-882-434a-2

Query Match
Best Local Similarity 6.2%; Score 30.8; DB 10; Length 493;
Pred. No. 7.8;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 219 AGGTATACATCAAGAGGAGGATGATCTTACACTACATGAGCAAACTGCTCTCTTA 278
DB 364 AGATACCTTTCAACCAAAAGGTTGATGCTCTGAGCTGAGCAAACTGCTGTTG 305
QY 279 CAACGAGCTGATGATGATGCTGCTGACACACAGTTGGTCCAGTCCAGGAGCATG 338

```

```

DB 304 CAACACACTGATCCAGCTGTTGTAGAGAGAGAGTTTGTCCAGTGTAGTGAAGT 245
QY 339 CAACCTTTTGGTGAAGATGATCT 364
DB 244 CATAGCTCCCTTCGATGATAGCT 219

RESULT 13
US-09-965-830-9
; Sequence 9, Application US/09965830
; Patent No. US20020177201A1
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: Y9903-PCT
; CURRENT APPLICATION NUMBER: US/09/965,830
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 09/600,776
; PRIOR FILING DATE: 2001-07-21
; PRIOR APPLICATION NUMBER: JP P1998-346198
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3715
; TYPE: DNA
; ORGANISM: Rattus sp.
US-09-965-830-9

Query Match
Best Local Similarity 6.2%; Score 30.8; DB 9; Length 3715;
Pred. No. 23;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 233 AGGAGGCTATGACTTCAGCTACAGTGAACAACTGCTCTCTACAAACAGCTGAT 292
DB 2772 AGTGAAGCAAGAAACAGACACTGAGCAAGCTACGAGCGGTGACGAGCTGCT 2831
QY 293 GCAGTGTGTTGACACACACCAG 314
DB 2832 GAACAGGTGCTGACAGTGCAG 2853

RESULT 14
US-09-929-230-3/c
; Sequence 3, Application US/09929230
; Patent No. US20020161203A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: RATTLESNAKE VENOM GLAND PROTEINS
; FILE REFERENCE: 00-72
; CURRENT APPLICATION NUMBER: US/09/929,230
; PRIOR FILING DATE: 2001-08-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
; NAME/KEY: misc feature
; LOCATION: 9, 12, 18, 21, 24, 27, 30, 36, 39, 42, 45, 48, 51, 54, 57,
; LOCATION: 66, 69, 78, 81, 105, 111, 117, 123, 135, 141, 153, 162,
; LOCATION: 168, 171, 174, 177, 180, 189, 192, 204, 210, 213, 219, 222,
; LOCATION: 225, 231, 234, 243, 246, 258, 261, 270, 297, 306, 312, 324
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 330, 333, 336, 339, 348, 351, 357, 360, 366, 372, 390, 393,
; LOCATION: 405, 408, 411, 417, 432, 435, 438, 441, 447, 450, 453
; OTHER INFORMATION: n = A,T,C or G
US-09-929-230-3

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2003, 07:11:59 ; Search time 1947 Seconds  
(without alignments)  
4100.860 Million cell updates/sec

Title: US-09-882-434A-2

Sequence: 1 attagctcttgagcttcac.....ccatgtgatgttacc 493

Scoring table: IDENTITY-NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_esba:\*  
2: em\_eschum:\*  
3: em\_escin:\*  
4: em\_escmu:\*  
5: em\_escov:\*  
6: em\_escpl:\*  
7: em\_escro:\*  
8: em\_escr:\*  
9: gb\_esc1:\*  
10: gb\_esc2:\*  
11: gb\_esc3:\*  
12: gb\_esc4:\*  
13: gb\_esc5:\*  
14: gb\_esc6:\*  
15: em\_esc7:\*  
16: em\_esc8:\*  
17: gb\_esc9:\*  
18: em\_esc10:\*  
19: em\_esc11:\*  
20: em\_esc12:\*  
21: em\_esc13:\*  
22: em\_esc14:\*  
23: em\_esc15:\*  
24: em\_esc16:\*  
25: em\_esc17:\*  
26: em\_esc18:\*  
27: em\_esc19:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.8	18.0	605	AW010330	ST04G06 P
2	88.8	18.0	726	B1416519	hasp001xd
3	86	17.4	750	B1416967	hasp002xp
4	85.6	17.2	598	AW043287	ST11E09 P
5	85	17.2	698	B1416872	hasp002xj
6	84.4	17.1	531	B1416595	hasp001xh

C	7	84.4	17.1	633	13	B1416764	B1416764	hasp002xd
C	8	84.4	17.1	639	13	B1416814	B1416814	hasp002xg
C	9	84.4	17.1	721	13	B1416901	B1416901	hasp002x1
C	10	81.8	16.6	530	13	B1416868	B1416868	hasp002xj
C	11	81.8	16.6	697	13	B1416667	B1416667	hasp001xn
C	12	81.2	16.5	531	13	B1416665	B1416665	hasp001xn
C	13	75.8	15.4	374	13	B1416726	B1416726	hasp001xn
C	14	57.4	11.6	476	13	B1416558	B1416558	hasp002xa
C	15	57	11.6	294	13	BE662406	ST62/ST62	BE662406
C	16	52.6	10.7	216	13	B1416830	hasp002xh	B1416830
C	17	48.6	9.9	825	13	B1416601	hasp001xi	B1416601
C	18	45.2	9.2	815	13	B1416504	hasp001xc	B1416504
C	19	42.6	8.6	620	12	BF299376	BF299376	BF299376
C	20	41.6	8.4	717	13	BM160808	EST563342	BM160808
C	21	39.6	8.0	592	13	AM257904	687063G11	AM257904
C	22	39.4	8.0	455	13	BM160535	EST563058	BM160535
C	23	39.4	8.0	504	13	B1473509	fp38b10. y	B1473509
C	24	39.4	8.0	660	13	BM163034	EST565557	BM163034
C	25	39	7.9	769	17	AQ751030	HS-5576. B	AQ751030
C	26	38.8	7.9	455	17	BH721895	BOHWT13TR	BH721895
C	27	38	7.7	576	17	AZ458762	IM0263F07	AZ458762
C	28	37.8	7.7	576	17	BE051991	GA_EA001	BE051991
C	29	37.4	7.6	589	12	BG136361	EST76803	BG136361
C	30	37.2	7.5	358	13	BM359106	CA_EA001	BM359106
C	31	37.2	7.5	977	17	CNS06031	TV_end of	AL407572
C	32	37	7.5	554	17	AZ012832	RPCI-23-3	AZ012832
C	33	37	7.5	558	12	BG733428	347312 MA	BG733428
C	34	36.8	7.5	539	17	AZ248820	RPCI-23-9	AZ248820
C	35	36.6	7.4	479	10	AW334443	S34H9 ACS	AW334443
C	36	36.6	7.4	996	9	AL568166	AL568166	AL568166
C	37	36.4	7.4	645	13	B1416777	hasp002xd	B1416777
C	38	36.2	7.3	774	17	BH422732	BOGRM52TF	BH422732
C	39	36.2	7.3	795	17	BH444988	BOGYA80TR	BH444988
C	40	36	7.3	265	12	BE981479	UI-M-CGP	BE981479
C	41	36	7.3	265	12	BE981840	UI-M-CGP	BE981840
C	42	36	7.3	265	12	BE981879	UI-M-CGP	BE981879
C	43	36	7.3	265	12	BE981926	UI-M-CGP	BE981926
C	44	35.8	7.3	392	10	AW417434	54088 MAR	AW417434
C	45	35.8	7.3	412	14	C82886	C82886	C82886

## ALIGNMENTS

RESULT 1  
LOCUS AW010330 605 bp mRNA linear EST 10-SEP-1999  
DEFINITION ST04G06 Pine Triplex shoot tip library pinus taeda cDNA clone  
ACCESSION ST04G06, mRNA sequence.  
VERSION AW010330  
KEYWORDS AW010330.1 GI:5859108  
SOURCE EST.  
ORGANISM loblolly pine.  
Pinus taeda

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
TITLE Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
JOURNAL 1 (bases 1 to 605)  
COMMENT Unpublished (1999)  
Contact: Ross Whetten  
Forest Biotechnology Group  
North Carolina State University  
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
NC, 27695-8008  
Tel: 919-515-7800  
Fax: 919-515-7801  
Email: rosswhetten@ncsu.edu  
Seq primer: 5' lambda Triplex2 Sequencing Primer.  
Location/Qualifiers  
1..605  
/organism="Pinus taeda"  
/db\_xref="taxon:3352"

## FEATURES

source

/clone="ST04C06"  
 /clone\_lib="Pine Triplex shoot tip library"  
 /lab\_host="E. coli BM25.8"  
 /notes="Organ: shoot tips; Vector: Lambda Triplex; Site\_1: SfiI (A); Site\_2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."  
 BASE COUNT 155 a 119 c 147 g 175 t 9 others  
 ORIGIN

Query Match 18.0%; Score 88.8; DB 10; Length 605;  
 Best Local Similarity 62.4%; Pred. No. 1.3e-14;  
 Matches 176; Conservative 0; Mismatches 97; Indels 9; Gaps 2;  
 QY 143 ATGGAGTGCAATTTACAGTATGGAGTGGTCCAGGTTGTAACAACCGTGTGAGCGATATA 202  
 Db 118 AGGCGAGTATTTTCACTCGCGGCGAGGGCTGTTGTAACAACCATGCTGCGCGATACA 177  
 QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256  
 Db 178 GCAATGTGGTGTCTTAATATCGCAACAACGTTTCATGGAGGATACGAGTTTCATGTATC 237  
 QY 257 CTGGACAAACTGCTGCTCTTACAAACAGGCTGGATGAGTGGTGTGGACACACACAGGT 316  
 Db 238 AAGCCAGACCGCTGCGCTTACAAACAGGACAACTGCAAGGGCGTTGCTCAAAACCGGT 297  
 QY 317 TTGGTCCAGTCCAG-----GCATGCAACCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373  
 Db 238 TTCTAGCAGTGTAGTCAAGCTTGCAGCAGTTTGGTTGGAAGAGTTTTCATCCAGT 357  
 QY 374 GCTAGATTTCAATCTCTTGATCCATCTTCTATGTTTCA 415  
 Db 358 GCTGAGTCTGTACAACTTTATATTACTATTATCACTAA 399

RESULT 2  
 BI416519 726 bp mRNA linear EST 15-AUG-2001  
 LOCUS  
 DEFINITION hasp001xd12f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xd12f, mRNA sequence.  
 ACCESSION BI416519 GI:15187542  
 VERSION BI416519  
 KEYWORDS EST.  
 SOURCE Pinus sylvestris/Heterobasidion annosum.  
 ORGANISM Eukaryota; mixed EST libraries.  
 REFERENCE 1 (bases 1 to 726)  
 ASIEGBU, F.O., Nahaikova, J., Choi, W., Stenlid, J. and Dean, R.A. Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris) unpublished (2001)  
 JOURNAL  
 COMMENT Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden  
 Tel: +46 18 67 15 98  
 Fax: +46 18 30 92 45  
 Email: Fred.Asiegbu@mykopat.slu.se  
 Seq primer: T7 primer.  
 Location/Qualifiers  
 1..726  
 /organism="Pinus sylvestris/Heterobasidion annosum"  
 /db\_xref="taxon:169015"  
 /clone="hasp001xd12f"  
 /clone\_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"  
 /dev\_stage="Seedling roots of scots pine were infected for

FEATURES  
 source

6 days with H. annosum"  
 /note="vector: pT-Adv; Site\_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."  
 BASE COUNT 170 a 148 c 193 g 211 t 4 others  
 ORIGIN

Query Match 18.0%; Score 88.8; DB 13; Length 726;  
 Best Local Similarity 63.1%; Pred. No. 1.4e-14;  
 Matches 173; Conservative 0; Mismatches 92; Indels 9; Gaps 2;  
 QY 143 ATGGAGTGGCATTTTACAGTATGGAGTGGTCCAGGTTGTAACAACCGTGTGAGCGATATA 202  
 Db 197 AGGCGAGTATTTTCACTCGCGTGGCAGGGCTGTTGTAACAACCATGCTGCGCGATACA 256  
 QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256  
 Db 257 GCAATGTGGTGTCTTAATATCGCAACAACGTTTCATGGAGGATACGAGTTTCATGTATC 316  
 QY 257 CTGGACAAACTGCTGCTCTTACAAACAGGCTGGATGAGTGGTGTGGACACACACAGGT 316  
 Db 317 AAGGCGAGACTGCTTCTGCTTACAAACAGGACAACTGCAAGGGCGTTGCTCAGACCCGGT 376  
 QY 317 TTGGTCCAGTCCCA---GGGATGCAACCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373  
 Db 377 TTCTAGCAGTGTATTAATCAAGCTTGCAGCAGTTTGGTTGGAAGAGTTTTCATCCAAT 436  
 QY 374 GCTAGATTTCAATCTCTTGATCCATCTTCTAT 407  
 Db 437 GCTGAGTGTGTGACAACTTTATATTACTATTAT 470

RESULT 3  
 BI416967/c 750 bp mRNA linear EST 15-AUG-2001  
 LOCUS  
 DEFINITION hasp002xpl5f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xpl5f, mRNA sequence.  
 ACCESSION BI416967 GI:15187990  
 VERSION BI416967  
 KEYWORDS EST.  
 SOURCE Pinus sylvestris/Heterobasidion annosum.  
 ORGANISM Eukaryota; mixed EST libraries.  
 REFERENCE 1 (bases 1 to 750)  
 ASIEGBU, F.O., Nahaikova, J., Choi, W., Stenlid, J. and Dean, R.A. Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris) unpublished (2001)  
 JOURNAL  
 COMMENT Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden  
 Tel: +46 18 67 15 98  
 Fax: +46 18 30 92 45  
 Email: Fred.Asiegbu@mykopat.slu.se  
 Seq primer: T7 primer.  
 Location/Qualifiers  
 1..750  
 /organism="Pinus sylvestris/Heterobasidion annosum"  
 /db\_xref="taxon:169015"  
 /clone="hasp002xpl5f"  
 /clone\_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"  
 /dev\_stage="Seedling roots of scots pine were infected for

6 days with H. annosum"  
 /note="vector: pT-Adv; Site\_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."  
 BASE COUNT 222 a 161 c 173 g 170 t 24 others  
 ORIGIN

Query Match 18.0%; Score 88.8; DB 13; Length 726;  
 Best Local Similarity 63.1%; Pred. No. 1.4e-14;  
 Matches 173; Conservative 0; Mismatches 92; Indels 9; Gaps 2;  
 QY 143 ATGGAGTGGCATTTTACAGTATGGAGTGGTCCAGGTTGTAACAACCGTGTGAGCGATATA 202  
 Db 197 AGGCGAGTATTTTCACTCGCGTGGCAGGGCTGTTGTAACAACCATGCTGCGCGATACA 256  
 QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256  
 Db 257 GCAATGTGGTGTCTTAATATCGCAACAACGTTTCATGGAGGATACGAGTTTCATGTATC 316  
 QY 257 CTGGACAAACTGCTGCTCTTACAAACAGGCTGGATGAGTGGTGTGGACACACACAGGT 316  
 Db 317 AAGGCGAGACTGCTTCTGCTTACAAACAGGACAACTGCAAGGGCGTTGCTCAGACCCGGT 376  
 QY 317 TTGGTCCAGTCCCA---GGGATGCAACCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373  
 Db 377 TTCTAGCAGTGTATTAATCAAGCTTGCAGCAGTTTGGTTGGAAGAGTTTTCATCCAAT 436  
 QY 374 GCTAGATTTCAATCTCTTGATCCATCTTCTAT 407  
 Db 437 GCTGAGTGTGTGACAACTTTATATTACTATTAT 470

FEATURES  
 source



## ORIGIN

Query Match 17.4%; Score 86; DB 13; Length 750;  
 Best Local Similarity 65.6%; Pred. No. 9.2e-14;  
 Matches 160; Conservative 0; Mismatches 75; Indels 9; Gaps 2;

143 ATGGAGTGCAATTACAGATGAGTGTCCAGGTTGTAAACACCGTGCAGCATATA 202  
 Db AGGGCAGTTATTACCTGCGGCGAGGCTGTGTAAACACAGTGGCGGATACA 182  
 628 AGGGCAGTTATTACCTGCGGCGAGGCTGTGTAAACACAGTGGCGGATACA 569  
 QY 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GAGGCTATGACTTCACTACA 256  
 Db 243 AGGGCAGTCCCTGCGGCTGTGTAAACACAGTGGCGGATACA 302  
 QY 317 TTGGGTCCAGTCCAG--GGCATCAACCTTTGGTGTGAAGATATCTTCAAT 373  
 Db 303 TTCTAGCAATGTATGATCAAGCTTCAGACAGTTGGTGAAGGTTTTCACACAGT 362  
 QY 374 GCTAATTCATACCTTGATCCATCTTATGTTTCA 415  
 Db 363 GCTAGTGTGTAAACATTTATATTACTATTATCACTAA 404

RESULT 4  
 AM043287 598 bp mRNA linear EST 18-SEP-1999  
 LOCUS ST31E09, Pine Triplex shoot tip library Pinus taeda cDNA clone

DEFINITION ST31E09, Pine Triplex shoot tip library Pinus taeda cDNA clone

ACCESSION AM043287  
 VERSION AM043287.1 GI:5903816

KEYWORDS EST.  
 SOURCE 10b101ly pine.

ORGANISM Pinus taeda  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.

REFERENCE 1 (bases 1 to 598) Retzel, E. and Sederoff, R.R.  
 Whetten, R.W., Kinlaw, C.S., The Pine Gene Discovery Project  
 Unpublished (1999)

JOURNAL Contact: Rose Whetten  
 Forest Biotechnology Group  
 North Carolina State University  
 Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
 NC, 27695-8008  
 Tel: 919-515-7800  
 Fax: 919-515-7801  
 Email: rosewhetten@unity.ncsu.edu

COMMENT Seq primer: 5' lambda Triplex2 Sequencing Primer.

FEATURES  
 Location/Qualifiers  
 1..598  
 /organism="Pinus taeda"  
 /db\_xref="taxon:3352"  
 /clone="ST31E09"  
 /clone\_lib="Pine Triplex shoot tip library"  
 /lab\_host="E. coli BM25.8"  
 /note="Organ: shoot tips; Vector: lambda Triplex; Site 1:  
 SfiI (A); Site 2: SfiI (B); Shoot tips (approx. 2 cm from  
 apex) were collected during the spring, frozen and used  
 for mRNA isolation. The SMART-PCR method (Clontech) was  
 used to prepare a library from 1 ug total RNA, using the  
 lambda Triplex vector. Plasmid subclones in pTriplex were  
 recovered by cre-lox excision in E. coli strain BM25.8 and  
 sequenced from the 5' end."

BASE COUNT 146 a 126 c 150 g 171 t 5 others  
 ORIGIN

Query Match 17.4%; Score 85.6; DB 10; Length 598;  
 Best Local Similarity 61.7%; Pred. No. 1.1e-13;  
 Matches 174; Conservative 0; Mismatches 99; Indels 9; Gaps 2;

QY 143 ATGGAGTGCAATTACAGATGAGTGTCCAGGTTGTAAACACCGTGCAGCATATA 202  
 Db 123 AGGGCAGTTATTACCTGCGGCGAGGCTGTGTAAACACAGTGGCGGATACA 182  
 QY 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GAGGCTATGACTTCACTACA 256  
 Db 183 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GAGGCTATGACTTCACTACA 242  
 QY 257 CTGCAAACTGCTGCTCTCTTACAAACGAGCTGATGAGTGTGTGACACACAGGT 316  
 Db 243 AGGGCAGTCCCTGCGGCTGTGTAAACACAGTGGCGGATACA 302  
 QY 317 TTGGGTCCAGTCCAG--GGCATCAACCTTTGGTGTGAAGATATCTTCAAT 373  
 Db 303 TTCTAGCAATGTATGATCAAGCTTCAGACAGTTGGTGAAGGTTTTCACACAGT 362  
 QY 374 GCTAATTCATACCTTGATCCATCTTATGTTTCA 415  
 Db 363 GCTAGTGTGTAAACATTTATATTACTATTATCACTAA 404

RESULT 5  
 B1416872 698 bp mRNA linear EST 15-AUG-2001  
 LOCUS hasp002xj10f Heterobasidion annosum - Scots pine infection stage  
 subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion  
 annosum cDNA clone hasp002xj10f, mRNA sequence.

DEFINITION B1416872, 1 GI:15187895

ACCESSION B1416872  
 VERSION B1416872.1 GI:15187895

KEYWORDS EST.  
 SOURCE Pinus sylvestris/Heterobasidion annosum.  
 ORGANISM Pinus sylvestris/Heterobasidion annosum  
 Eukaryota; mixed EST libraries.

REFERENCE 1 (bases 1 to 698) Aiegbu, F.O., Nahlikova, J., Choi, W., Stenlid, J. and Dean, R.A.  
 Expressed sequence tags of randomly selected cDNA clones from the  
 interaction of the root rot fungus (Heterobasidion annosum) with  
 seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)  
 Contact: Fred O. Aiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,  
 Sweden

COMMENT Tel: +46 18 67 15 98  
 Fax: +46 18 30 92 45  
 Email: Fred.Aiegbu@mykopat.slu.se  
 Seq primer: T7 primer.

FEATURES  
 Location/Qualifiers  
 1..698  
 /organism="Pinus sylvestris/Heterobasidion annosum"  
 /db\_xref="taxon:168015"  
 /clone="hasp002xj10f"  
 /clone\_lib="Heterobasidion annosum - Scots pine infection  
 stage subtraction cDNA library (hasp)"  
 /dev\_stage="Seedling roots of scots pine were infected for  
 6 days with H. annosum"  
 /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive  
 hybridization cDNA library was constructed from scots  
 pine roots infected for 6-days with mycelia of  
 Heterobasidion annosum (FPS)."

BASE COUNT 174 a 143 c 173 g 208 t

ORIGIN

Query Match 17.2%; Score 85; DB 13; Length 698;  
 Best Local Similarity 65.4%; Pred. No. 1.7e-13;  
 Matches 159; Conservative 0; Mismatches 75; Indels 9; Gaps 2;  
 QY 143 ATGGAGTGCAATTACAGATGAGTGTCCAGGTTGTAAACACCGTGCAGCATATA 202  
 Db 144 AGGGCAGTTATTACCTGCGGCGAGGCTGTGTAAACACAGTGGCGGATACA 203  
 QY 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GAGGCTATGACTTCACTACA 256

```

Db 204 GCAATGTGGGTGTTCTAATATCGCAACAACGTTCTATGAGGATACAGATTCTATGATC 263
QY 257 CTGGCAAACTGTGCTCTCTACAAACGAGGCTGGATGACAGTGTGTTGCACACACCAAGT 316
Db 264 AAGGCCAGACCGCTGCGGCTTACAACAACGGAACACTGCAAGGGGTTGCTCAGACCCGCT 323
QY 317 TTGGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373
Db 324 TTTCAGCAGTGTAAATCAAGCTTGCAGCAGTTTGGTTGGAAGAGTTTTTTCATCCAGT 383
QY 374 GCT 376
Db 384 GCT 386

RESULT 6
BI416595 531 bp mRNA linear EST 15-AUG-2001
LOCUS hasp001xh23f Heterobasidion annosum - Scots pine infection stage
DEFINITION subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp001xh23f, mRNA sequence.
ACCESSION BI416595
VERSION BI416595.1 GI:15187618
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 531)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
    source
        1..531
            /organism="Pinus sylvestris/Heterobasidion annosum"
            /db_xref="taxon:169015"
            /clone="hasp001xh23f"
            /clone_lib="Heterobasidion annosum - Scots pine infection
            stage subtraction cDNA library (hasp)"
            /dev_stages="Seedling roots of scots pine were infected for
            6 days with H. annosum"
            /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
            hybridization cDNA library was constructed from scots
            pine roots infected for 6-days with mycelia of
            Heterobasidion annosum (FP5)."
BASE COUNT 115 a 117 c 138 g 161 t
ORIGIN
    Query Match 17.1%; Score 84.4; DB 13; Length 531;
    Best Local Similarity 65.2%; Pred. No. 2.4e-13;
    Matches 159; Conservative 0; Mismatches 16; Indels 9; Gaps 2;

QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTTGTAAACACCGTGTGAGCGATATA 202
Db 141 AGGGCAGTTATTTCACTCGTGGGCGAGGCGCTGGTTGCAACCACTGCTCGCTGATACA 200
QY 203 GCAAGTGTGGATGCTCAGCTATACATACAGAG-----GGAGGCTATGACTTCAGCTACA 256
Db 201 GCAATGTGGGTGTTCTAATATCGGTACGACGTTCTATGAGGATACGAGTTCTGTGATC 260
QY 257 CTGGCAAACTGCTGCTCTCTACAAACGAGGCTGGATGCACTGGTGTGTCACACACCAAGT 316
Db 261 AAGGCCAGACTGCTTCTGCTTACAACAACGCGCAACTGCAAGGCGTTGCTCAGACCCGTT 320

```

```

QY 317 TTGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373
Db 321 TTTCGCAAGTGTATCAAGATTGCGAGCGGTTTGGTTGGAAGAGTTTCTTCATCCAGT 380
QY 374 GCTA 377
Db 381 GCTA 384

RESULT 7
BI416764/c 633 bp mRNA linear EST 15-AUG-2001
LOCUS hasp002xd04f Heterobasidion annosum - Scots pine infection stage
DEFINITION subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp002xd04f, mRNA sequence.
ACCESSION BI416764
VERSION BI416764.1 GI:15187787
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 633)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
    source
        1..633
            /organism="Pinus sylvestris/Heterobasidion annosum"
            /db_xref="taxon:169015"
            /clone="hasp002xd04f"
            /clone_lib="Heterobasidion annosum - Scots pine infection
            stage subtraction cDNA library (hasp)"
            /dev_stages="Seedling roots of scots pine were infected for
            6 days with H. annosum"
            /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
            hybridization cDNA library was constructed from scots
            pine roots infected for 6-days with mycelia of
            Heterobasidion annosum (FP5)."
BASE COUNT 197 a 154 c 127 g 155 t
ORIGIN
    Query Match 17.1%; Score 84.4; DB 13; Length 633;
    Best Local Similarity 65.2%; Pred. No. 2.5e-13;
    Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTTGTAAACACCGTGTGAGCGATATA 202
Db 540 AGGGCAGTTATTTCACTCGTGGGCGAGGCGCTGGTTGTAACCACTGCTGCGCGATACA 481
QY 203 GCAAGTGTGGATGCTCAGCTATACATACAGAGG-----AGGCTATGACTTCAGCTACA 256
Db 480 GCAATGTGGGTGTTCTAATATCGGCAACAACGTTCTATGAGGATACGAGTTCTGTGATC 421
QY 257 CTGGCAAACTGCTGCTCTCTACAAACGAGGCTGGATGCACTGGTGTGTCACACACCAAGT 316
Db 420 AAGGCCAGACTGCTTCTGCTTACAACAACGCGCAACTGCAAGGCGTTGCTCAGACCCGTT 361
QY 317 TTGGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373
Db 360 TTTCGCAAGTGTAAATCAAGCTTGCAGCGGTTTGGTTGGAAGAGTTTTTTCATCCAGT 301
QY 374 GCTA 377

```

DB 300 GCTA 297

RESULT 8  
LOCUS B1416814/c 639 bp mRNA linear EST 15-AUG-2001

DEFINITION hasp002xg03f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xg03f, mRNA sequence.

ACCESSION B1416814

VERSION B1416814

KEYWORDS B1416814.1 GI:15187837

SOURCE EST.

ORGANISM Pinus sylvestris/Heterobasidion annosum.

REFERENCE 1 (bases 1 to 639)  
Asiegbu, F.O., Nahlikova, J., Choi, W., Stenlid, J. and Dean, R.A. Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris) Unpublished (2001)

JOURNAL Contact: Fred O. Asiegbu  
Dept. of Forest Mycology & Pathology  
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden  
Tel: +46 18 67 15 98  
Fax: +46 18 30 92 45  
Email: Fred.Asiegbu@mykopat.slu.se

COMMENT Seq primer: T7 primer

FEATURES  
source  
1. .639  
/organism="Pinus sylvestris/Heterobasidion annosum"  
/db\_xref="taxon:169015"  
/clone="hasp002xg03f"  
/clone\_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"  
/dev\_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"  
/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6 days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 199 a 154 c 127 g 159 t

ORIGIN

Query Match 17.1%; Score 84.4; DB 13; Length 639;  
Best Local Similarity 65.2%; Pred. No. 2.5e-13;  
Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

DB 143 ATGGAGTGCATTTCAGATGAGTGTCCAGTTGTAAACACCTGTGAGCATATA 202  
DB 547 AGGCGAGTATTTCACGTCGGGCGAGGCGCTGTGTAAACACCTGTGAGCATATA 488  
DB 203 GCAAGTGTGATGCTCAGCTATACATCAAGAAAGG-----AGCTATGACTTCACTACA 256  
DB 487 GCAATGTGGGTGTTCTAATATCGGCAACAGTTTCATGACGAGTACGAGTTCGTATC 428  
DB 257 CTGACAACTGCTGCTCTACAAACGAGGTGATGAGTGTGACACACACAGGT 316  
DB 427 AAGGCCAGACTGCTTCTGCTTACAAACGCGCAACGTAAGAGGCGTGTGACAGCCGTT 368  
DB 317 TTGGGTCCAGTGCCA---GGGCAATGCAACCTTTTGGTTGGAAGATATCTTCATCCAA 373  
DB 367 TTTCTGGCAGTGTAAATCAAGCTTGACAGCGGTTTGGTTGGAAGATTTTTCATCCAGT 308  
DB 374 GCTA 377  
DB 307 GCTA 304

RESULT 9  
B1416901  
B1416901

LOCUS B1416901 721 bp mRNA linear EST 15-AUG-2001

DEFINITION hasp002xj03f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xj03f, mRNA sequence.

ACCESSION B1416901

VERSION B1416901

KEYWORDS B1416901.1 GI:15187924

SOURCE EST.

ORGANISM Pinus sylvestris/Heterobasidion annosum.

REFERENCE 1 (bases 1 to 721)  
Asiegbu, F.O., Nahlikova, J., Choi, W., Stenlid, J. and Dean, R.A. Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris) Unpublished (2001)

JOURNAL Contact: Fred O. Asiegbu  
Dept. of Forest Mycology & Pathology  
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden  
Tel: +46 18 67 15 98  
Fax: +46 18 30 92 45  
Email: Fred.Asiegbu@mykopat.slu.se

COMMENT Seq primer: T7 primer

FEATURES  
source  
1. .721  
/organism="Pinus sylvestris/Heterobasidion annosum"  
/db\_xref="taxon:169015"  
/clone="hasp002xj03f"  
/clone\_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"  
/dev\_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"  
/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6 days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 165 a 146 c 177 g 227 t 6 others

ORIGIN

Query Match 17.1%; Score 84.4; DB 13; Length 721;  
Best Local Similarity 65.2%; Pred. No. 2.6e-13;  
Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

DB 143 ATGGAGTGCATTTCAGATGAGTGTCCAGTTGTAAACACCTGTGAGCATATA 202  
DB 213 AGGCGAGTATTTCACGTCGGGCGAGGCGCTGTGTAAACACCTGTGAGCATATA 272  
DB 203 GCAAGTGTGATGCTCAGCTATACATCAAGAAAGG-----AGCTATGACTTCACTACA 256  
DB 273 GCAATGTGGGTGTTCTAATATCGGCAACAGTTTCATGACGAGTACGAGTTCGTATC 332  
DB 257 CTGACAACTGCTGCTCTACAAACGAGGTGATGAGTGTGACACACACAGGT 316  
DB 333 AAGGCCAGACTGCTTCTGCTTACAAACGCGCAACGTAAGAGGCGTGTGACAGCCGTT 392  
DB 317 TTGGGTCCAGTGCCA---GGGCAATGCAACCTTTTGGTTGGAAGATATCTTCATCCAA 373  
DB 393 TTTCTGGCAGTGTAAATCAAGCTTGACAGCGGTTTGGTTGGAAGATTTTTCATCCAGT 452  
DB 374 GCTA 377  
DB 453 GCTA 456

RESULT 10  
B1416868  
B1416868/c 530 bp mRNA linear EST 15-AUG-2001

LOCUS B1416868

DEFINITION hasp002xj06f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xj06f, mRNA sequence.

ACCESSION B1416868

VERSION B1416868.1 GI:15187891

KEYWORDS EST.  
SOURCE Pinus sylvestris/Heterobasidion annosum.  
ORGANISM Pinus sylvestris/Heterobasidion annosum  
REFERENCE Eukaryota; mixed EST libraries.  
AUTHORS 1 (bases 1 to 530)  
TITLE Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.  
Expressed sequence tags of randomly selected cDNA clones from the  
interaction of the root rot fungus (Heterobasidion annosum) with  
seedling roots of Scots pine (Pinus sylvestris)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Fred O. Asiegbu  
Dept. of Forest Mycology & Pathology  
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,  
Sweden  
Tel: +46 18 67 15 98  
Fax: +46 18 30 92 45  
Email: Fred.Asiegbu@mykopat.slu.se  
Seq primer: T7 primer.  
Location/Qualifiers  
FEATURES  
source  
1..530  
/organism="Pinus sylvestris/Heterobasidion annosum"  
/db\_xref="taxon:169015"  
/clone="hasp002xj06f"  
/clone\_lib="Heterobasidion annosum - Scots pine infection  
stage subtraction cDNA library (hasp)"  
/dev\_stages="Seedling roots of scots pine were infected for  
6 days with H. annosum"  
/note="Vector: pT-Adv; Site 1: EcoRI; The subtractive  
hybridization cDNA library was constructed from scots  
pine roots infected for 6-days with mycelia of  
Heterobasidion annosum (FPS)."  
BASE COUNT 142 a 124 c 107 g 157 t  
ORIGIN  
Query Match 16.6%; Score 81.8; DB 13; Length 530;  
Best Local Similarity 64.6%; Pred. No. 1.3e-12;  
Matches 157; Conservative 0; Mismatches 77; Indels 9; Gaps 2;  
QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTTGTAACAACCGTCTGAGCGATATA 202  
Db 404 AGGCGAGTATTTCACCTGCTGGCGAGGCGCTGTTGTAACAACCATGCTCGCGATACA 345  
QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256  
Db 344 GCAATATGGGGTGTAAATATCGGCAACACAGTTCATGGAGGATACAGATTCATGTATC 285  
QY 257 CTGGACAAACTGCTCTCTCAACACAGGCTGGATGAGTGGTGTGCACACACACAGGT 316  
Db 284 AGGCGCAGACCGCTCGGCTTACACACGACACTGCAAGGGCGTGTCTCAGACCCGGT 225  
QY 317 TTGGTCCAGTGCCA---GGGATGCAACCCCTTTGGTTGGAAGAGTATCTTCATCCAAT 373  
Db 224 TTTTAGCAGTGTAAATCAAGCTGCACAGTTTGGTTGGAAAAGTTTTTCATCCAGT 165  
QY 374 GCT 376  
Db 164 GCT 162  
RESULT 11  
BI416667  
LOCUS BI416667 597 bp mRNA linear EST 15-AUG-2001  
DEFINITION hasp001xn07f Heterobasidion annosum - Scots pine infection stage  
subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion  
annosum cDNA clone hasp001xn07f, mRNA sequence.  
ACCESSION BI416667  
VERSION BI416667.1 GI:15187690  
KEYWORDS EST.  
SOURCE Pinus sylvestris/Heterobasidion annosum.  
ORGANISM Pinus sylvestris/Heterobasidion annosum  
REFERENCE Eukaryota; mixed EST libraries.  
AUTHORS 1 (bases 1 to 697)  
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

TITLE Expressed sequence tags of randomly selected cDNA clones from the  
interaction of the root rot fungus (Heterobasidion annosum) with  
seedling roots of Scots pine (Pinus sylvestris)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Fred O. Asiegbu  
Dept. of Forest Mycology & Pathology  
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,  
Sweden  
Tel: +46 18 67 15 98  
Fax: +46 18 30 92 45  
Email: Fred.Asiegbu@mykopat.slu.se  
Seq primer: T7 primer.  
Location/Qualifiers  
FEATURES  
source  
1..697  
/organism="Pinus sylvestris/Heterobasidion annosum"  
/db\_xref="taxon:169015"  
/clone="hasp001xn07f"  
/clone\_lib="Heterobasidion annosum - Scots pine infection  
stage subtraction cDNA library (hasp)"  
/dev\_stages="Seedling roots of scots pine were infected for  
6 days with H. annosum"  
/note="Vector: pT-Adv; Site 1: EcoRI; The subtractive  
hybridization cDNA library was constructed from scots  
pine roots infected for 6-days with mycelia of  
Heterobasidion annosum (FPS)."  
BASE COUNT 155 a 145 c 175 g 222 t  
ORIGIN  
Query Match 16.6%; Score 81.8; DB 13; Length 697;  
Best Local Similarity 64.6%; Pred. No. 1.4e-12;  
Matches 157; Conservative 0; Mismatches 77; Indels 9; Gaps 2;  
QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTTGTAACAACCGTCTGAGCGATATA 202  
Db 143 AGGCGAGTATTTCACCTGCTGGCGAGGCGCTGTTGTAACAACCATGCTCGACGATACA 202  
QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256  
Db 203 GCAATATGGGGTGTCTAATATCGGCAACACAGTTCATGGAGGATACAGTTCATGTATC 262  
QY 257 CTGGACAAACTGCTCTCTCAACACAGGCTGGATGAGTGGTGTGCACACACACAGGT 316  
Db 263 AAGGCCAGACCGCTCGGCTTACACACGACACTGCAAGGGCGTGTCTCATACCCGGT 322  
QY 317 TTGGTCCAGTGCCA---GGGATGCAACCCCTTTGGTTGGAAGAGTATCTTCATCCAAT 373  
Db 323 TTTTAGCAGTGTAAATCAAGCTGCACAGTTTGGTTGGAAAAGTTTTTCATCCAGT 382  
QY 374 GCT 376  
Db 383 GCT 385  
RESULT 12  
BI416665  
LOCUS BI416665 531 bp mRNA linear EST 15-AUG-2001  
DEFINITION hasp001xn05f Heterobasidion annosum - Scots pine infection stage  
subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion  
annosum cDNA clone hasp001xn05f, mRNA sequence.  
ACCESSION BI416665  
VERSION BI416665.1 GI:15187688  
KEYWORDS EST.  
SOURCE Pinus sylvestris/Heterobasidion annosum.  
ORGANISM Pinus sylvestris/Heterobasidion annosum  
REFERENCE Eukaryota; mixed EST libraries.  
AUTHORS 1 (bases 1 to 531)  
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.  
Expressed sequence tags of randomly selected cDNA clones from the  
interaction of the root rot fungus (Heterobasidion annosum) with  
seedling roots of Scots pine (Pinus sylvestris)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Fred O. Asiegbu  
Dept. of Forest Mycology & Pathology

Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden  
 Tel: +46 18 67 15 98  
 Fax: +46 18 30 92 45  
 Email: Fred.Asiegbu@mykopac.slu.se  
 Seq primer: T7 primer.

## FEATURES

source

Location/Qualifiers

1..531  
 /organism="Pinus sylvestris/Heterobasidion annosum"  
 /db\_xref="taxon:169015"  
 /clone="hasp001x05f"  
 /clone\_1ib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"  
 /dev\_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"  
 /note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 117 a 118 c 138 g 158 t  
 ORIGIN

Query Match 16.5%; Score 81.2; DB 13; Length 531;  
 Best Local Similarity 64.3%; Pred. No. 2e-12; Indels 9; Gaps 2;  
 Matches 157; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

QY 143 ATGGAGTGCAATTACAGATGAGTGCTCCAGTTGTAAACACCGTGTGACGATATA 202  
 DB 136 AGGCGAGTATTTCACGTGCGGCGAGGCGCTGTGTAAACACCATGCTGCGGATACA 195  
 QY 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCAGTTACA 256  
 DB 196 GCAAGTGTGATGCTTAAATATCGTACACACCTCATGAGGATACGAGTTGCTATATC 255  
 QY 257 CTGGACAACCTGCTGCTCTACACACGCTGATGCTGTGTGACACACAGAGT 316  
 DB 256 AAGGCCAGACCGCTGCGGCTTACACACGCGCAACTGCAAGGCGTGTCTCATATCCCGT 315  
 QY 317 TTGGGTCAGTGCCA--GGGCAATGCAACCCCTTTGGTGAAGATATCTTCATCAAT 373  
 DB 316 TTTCGTGCGAGTTTAAATCAAGCTTGACCGGTTTGGTGGAAAAATTTTTCATCACT 375  
 QY 374 GCTA 377  
 DB 376 GCTA 379

## RESULT 13

B1416726

374 bp mRNA linear EST 15-AUG-2001

## LOCUS

hasp002xa22f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xa22f, mRNA sequence.

## ACCESSION

B1416726

## VERSION

B1416726.1 GI:15187749

## KEYWORDS

Pinus sylvestris/Heterobasidion annosum.

## SOURCE

Pinus sylvestris/Heterobasidion annosum.

## ORGANISM

Eukaryota; mixed EST libraries.

## REFERENCE

1 (bases 1 to 374)  
 Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

## AUTHORS

Expressed sequence tags of randomly selected cDNA clones from the

## TITLE

interaction of the root rot fungus (Heterobasidion annosum) with

## JOURNAL

seedling roots of Scots pine (Pinus sylvestris)

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## FEATURES

source

Location/Qualifiers

1..374  
 /organism="Pinus sylvestris/Heterobasidion annosum"  
 /db\_xref="taxon:169015"  
 /clone="hasp002xa22f"  
 /clone\_1ib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"  
 /dev\_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"  
 /note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 83 a 85 c 104 g 102 t  
 ORIGIN

Query Match 15.4%; Score 75.8; DB 13; Length 374;  
 Best Local Similarity 64.6%; Pred. No. 6.2e-11; Indels 9; Gaps 2;  
 Matches 148; Conservative 0; Mismatches 72; Indels 9; Gaps 2;

QY 143 ATGGAGTGCAATTACAGATGAGTGCTCCAGTTGTAAACACCGTGTGACGATATA 202  
 DB 146 AGGCGAGTATTTCACGTGCGGCGAGGCGCTGTGTAAACACCATGCTGCGGATACA 205  
 QY 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCAGTTACA 256  
 DB 206 GCAATGTGGGTGTTTAAATATCGGCAACACCTTCATGAGGATACGAGTTGCTATATC 265  
 QY 257 CTGGACAACCTGCTGCTCTACACACGCTGATGCTGTGTGACACACAGAGT 316  
 DB 266 AAGGCCAGACCTGCTGCTTACACACGAGCAACTGCAAGGCGTGTGCTACACCGGT 325  
 QY 317 TTGGGTCAGTGCCA--GGGCAATGCAACCCCTTTGGTGAAGATAT 362  
 DB 326 TTTCAGAGTATTAAATCAAGCTTGACCGGTTTGGTGGAAAAATTTTTCATCACT 374

## RESULT 14

B1416558

476 bp mRNA linear EST 15-AUG-2001

## LOCUS

hasp001xf17f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xf17f, mRNA sequence.

## ACCESSION

B1416558

## VERSION

B1416558.1 GI:15187581

## KEYWORDS

Pinus sylvestris/Heterobasidion annosum.

## SOURCE

Pinus sylvestris/Heterobasidion annosum.

## ORGANISM

Eukaryota; mixed EST libraries.

## REFERENCE

1 (bases 1 to 476)  
 Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

## AUTHORS

Expressed sequence tags of randomly selected cDNA clones from the

## TITLE

interaction of the root rot fungus (Heterobasidion annosum) with

## JOURNAL

seedling roots of Scots pine (Pinus sylvestris)

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## JOURNAL

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

## COMMENT

Unpublished (2001)  
 Contact: Fred O. Asiegbu  
 Dept. of Forest Mycology & Pathology  
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

hybridization cDNA library was constructed from Scots pine roots infected for 6-days with mycelia of *Heterobasidion annosum* (FPS)."

BASE COUNT 141 a 114 c 91 g 129 t 1 others  
 ORIGIN  
 Query Match 11.6%; Score 57.4; DB 13; Length 476;  
 Best Local Similarity 66.7%; Pred. No. 1.5e-05;  
 Matches 98; Conservative 0; Mismatches 46; Indels 3; Gaps 1;

QY 233 AGGAGGCTATGACTTACAGTACACTGACAACTGCTCTCTACAAACAGGCTGGAT 292  
 |||||  
 Db 452 ATGGAGGATACAGTTCTATGATCAAGCCAGACGGCTGCGGCTTACAAACGACAACT 393  
 |||||  
 QY 293 GCAGTGGTGTGCACACACACAGGTTGGGTCAGTGCAGG---GCATGCAACCCCTTTTG 349  
 |||||  
 Db 392 GCAAGGGGTTGCTCAGACCCGGTTTCTAGCAGTGTGTAATCAAGCTTGACGACGATTTTG 333  
 |||||  
 QY 350 GTTGAAGAGTATCTTCATCCAAATGCT 376  
 |||||  
 Db 332 GTTGAAGAGTTTTTTTCATCCAGTCT 306  
 |||||

RESULT 15  
 BE662406 294 bp mRNA linear EST 08-SEP-2000  
 DEFINITION ST62/ST62H06 Pine Triplex shoot tip library Pinus taeda cDNA clone  
 ST62/ST62H06, mRNA sequence.  
 ACCESSION BE662406  
 VERSION BE662406.1 GI:9992999  
 KEYWORDS EST.  
 SOURCE loblolly pine.  
 ORGANISM Pinus taeda  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 1 (bases 1 to 294)  
 AUTHORS Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.  
 TITLE The Pine Gene Discovery Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Ross Whetten  
 Forest Biotechnology Group  
 North Carolina State University  
 Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
 NC, 27695-8008  
 Tel: 919-515-7800  
 Fax: 919-515-7801  
 Email: rosswhet@unity.ncsu.edu  
 Seq primer: 5' lambda Triplex2 Sequencing Primer.  
 Location/Qualifiers  
 1..294  
 /organism="Pinus taeda"  
 /db\_xref="taxon:3352"  
 /clone="ST62/ST62H06"  
 /lab\_host="E. coli BM25.8"  
 /note="Organ: shoot tips; Vector: Lambda Triplex; Site: 1:  
 Sfil (A); Site 2: Sfil (B); Shoot tips (approx. 2 cm from  
 apex) were collected during the spring, frozen and used  
 for mRNA isolation. The SMART-PCR method (Clontech) was  
 used to prepare a library from 1 ug total RNA, using the  
 Lambda Triplex vector. Plasmid subclones in pTriplex were  
 recovered by cre-lox excision in E. coli strain BM25.8 and  
 sequenced from the 5' end."

BASE COUNT 70 a 67 c 69 g 77 t 11 others  
 ORIGIN  
 Query Match 11.6%; Score 57; DB 10; Length 294;  
 Best Local Similarity 60.2%; Pred. No. 1.5e-05;  
 Matches 106; Conservative 0; Mismatches 64; Indels 6; Gaps 1;

QY 143 ATGGAGTGCAATTTACAGTATGGAGTGTCCAGGTTGTAAACACCGCTGCGGATATA 202  
 |||||  
 Db 118 AGGGCANTTATTTCACTCGGTGGGAGGCGCTNNNTTGTAAACACCATGCTGCGGATACA 177  
 |||||

QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256  
 |||||  
 Db 178 GCAAAATGTGGTGTCTTAATATCGGCAACAACGTTTCATGGAGGATACGAGTTTCATGTATC 237  
 |||||  
 QY 257 CTGGCAAACTGCTGCTCTCTACAAACAGGCTGGGATGCAAGTGGTGTGTCACACACACC 312  
 |||||  
 Db 238 AAGNNCAGACCNNNTGCGNNNTTACAACACGGACAACCTGCAAGGGCGTGTCTCAAACC 293  
 |||||

Search completed: January 12, 2003, 08:35:41  
 Job time : 1953 secs